

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model
Run on: April 3, 2004, 16:47:19 ; Search time 3552.46 Seconds
(without alignments)
5295.174 Million cell updates/sec

Title: US-09-985-689A-2
Perfect score: 2245
Sequence: 1 NDVARGIVKADVAQSYGLY.....EVQAYNVFPGQNFSLAIVN 434

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n_model -DSV=xlh
-Q=/cgn2_1/USPTO.spool/US0985689/runat_31032004_161806_4155/app_query.fasta_1.3498
-DS=genEmbl -QMM=fastap -SUFFIX=rgc -MINMATCH=0.1 -DOOPCL=0 -DOOEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-CUTFM=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US0985689@cgn 1 11925 @runat_31032004_161806_4155 -NCPU=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOC=100 -LORGLQ
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.*
1: gb_ba.*
2: gb_htg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*
15: gb_wa.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*

29: em_vi.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htg_mus.*
34: em_htg_pln.*
35: em_htg_rtd.*
36: em_htg_nam.*
37: em_htg_vrt.*
38: em_sy.*
39: em_hgo_hum.*
40: em_hgo_mus.*
41: em_hgo_other.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result | No. | Score | Match | Length | DB | ID | Description |
|--------|--------|-------|--------|--------|-----------|-----------|--------------|
| 1 | 2237 | 99.6 | 1920 | 1 | AB046403 | AB046403 | Bacillus |
| 2 | 2237 | 99.6 | 1920 | 6 | AB368116 | AB368116 | Sequence |
| 3 | 2197 | 97.9 | 1823 | 6 | AB368118 | AB368118 | Sequence |
| 4 | 2196 | 97.8 | 1823 | 1 | AB084155 | AB084155 | Bacillus |
| 5 | 2191 | 97.6 | 1305 | 6 | AX839476 | AX839476 | Sequence |
| 6 | 2191 | 97.6 | 1823 | 1 | AB051423 | AB051423 | Bacillus |
| 7 | 2191 | 97.6 | 1823 | 6 | AR368117 | AR368117 | Sequence |
| 8 | 2130 | 94.9 | 1302 | 1 | AB046406 | AB046406 | Bacillus |
| 9 | 2118.5 | 94.4 | 3003 | 6 | AR069954 | AR069954 | Sequence |
| 10 | 2118.5 | 94.4 | 3003 | 6 | BD062155 | BD062155 | Nucleic a |
| 11 | 2010.5 | 89.6 | 1299 | 1 | AB046405 | AB046405 | Bacillus |
| 12 | 2006.5 | 89.4 | 1299 | 1 | AB046402 | AB046402 | Bacillus |
| 13 | 1999.5 | 89.1 | 1299 | 1 | AB046404 | AB046404 | Bacillus |
| 14 | 1998.5 | 89.0 | 2218 | 6 | EO3808 | EO3808 | DNA encodin |
| 15 | 626.5 | 27.9 | 60006 | 1 | AF268611 | AF268611 | Unculture |
| 16 | 571 | 25.4 | 301205 | 1 | AE017218 | AE017218 | Geobacter |
| 17 | 526.5 | 23.5 | 5890 | 3 | DDU60086 | DDU60086 | Dicystosteli |
| 18 | 509.5 | 22.7 | 5772 | 3 | AF466309 | AF466309 | Dicystosteli |
| 19 | 490 | 21.8 | 6115 | 3 | DDU20432 | DDU20432 | Dicystosteli |
| 20 | 457.5 | 20.4 | 1977 | 6 | AR201152 | AR201152 | Sequence |
| 21 | 443 | 19.7 | 5109 | 3 | AF263455 | AF263455 | Dicystosteli |
| 22 | 428.5 | 19.1 | 126928 | 2 | AC096673 | AC096673 | Trypanoso |
| 23 | 425.5 | 19.0 | 1236 | 6 | AR201146 | AR201146 | Sequence |
| 24 | 425.5 | 19.0 | 1962 | 6 | AR201155 | AR201155 | Sequence |
| 25 | 425.5 | 19.0 | 12452 | 1 | AE010265 | AE010265 | Pyrococcus |
| 26 | 418 | 18.6 | 14376 | 1 | AE013049 | AE013049 | Thermoana |
| 27 | 407.5 | 18.2 | 303450 | 1 | SC0938130 | SC0938130 | Streptomy |
| 28 | 387 | 17.2 | 299850 | 1 | AP004601 | AP004601 | Oceanobac |
| 29 | 365 | 16.3 | 3348 | 1 | D83672 | D83672 | Streptomyce |
| 30 | 360.5 | 16.1 | 300425 | 1 | AF005044 | AF005044 | Streptomy |
| 31 | 359.5 | 16.0 | 2539 | 6 | AR202322 | AR202322 | Sequence |
| 32 | 357.5 | 15.9 | 292550 | 1 | AP001513 | AP001513 | Bacillus |
| 33 | 354.5 | 15.8 | 2809 | 6 | AR202321 | AR202321 | Sequence |
| 34 | 354.5 | 15.8 | 3900 | 1 | AB007809 | AB007809 | Streptomy |
| 35 | 351 | 15.6 | 291000 | 1 | SC0939105 | SC0939105 | Streptomy |
| 36 | 345.5 | 15.4 | 302500 | 1 | AP005034 | AP005034 | Streptomy |
| 37 | 343 | 15.3 | 135638 | 1 | AF484556 | AF484556 | Streptomy |
| 38 | 340 | 15.1 | 1239 | 1 | AF305633 | AF305633 | Thermoana |
| 39 | 340 | 15.1 | 1239 | 1 | AY028704 | AY028704 | Thermoana |
| 40 | 339 | 15.1 | 12395 | 1 | AE013026 | AE013026 | Thermoana |
| 41 | 338.5 | 15.1 | 1329 | 6 | AX433519 | AX433519 | Sequence |
| 42 | 331.5 | 14.8 | 301150 | 1 | AP004602 | AP004602 | Oceanobac |
| 43 | 331 | 14.7 | 207829 | 1 | BSUB0010 | BSUB0010 | Bacillus su |
| 44 | 329.5 | 14.7 | 300800 | 1 | SC0938112 | SC0938112 | Streptomy |
| 45 | 323.5 | 14.4 | 2731 | 1 | AF015225 | AF015225 | Thermococ |

ALIGNMENTS

RESULT 1

AB046403
 LOCUS AB046403 1920 bp DNA linear BCT 10-MAY-2002
 DEFINITION Bacillus sp. 9860 PROA gene for protease, complete cds.
 ACCESSION AB046403
 VERSION AB046403.2 GI:20521152
 KEYWORDS
 SOURCE Bacillus sp. 9860
 ORGANISM Bacillus sp. 9860
 Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 1
 AUTHORS Saeki K., Okuda M., Hatada Y., Kobayashi T., Ito S., Takami H. and
 Horikoshi K.
 TITLE Novel oxidatively stable subtilisin-like serine proteases from
 alkaliphilic Bacillus spp.: enzymatic properties, sequences, and
 evolutionary relationships
 JOURNAL Biochem. Biophys. Res. Commun. 279 (2), 313-319 (2000)
 MEDLINE 20568675
 PUBMED 11118284
 REFERENCE 2 (bases 1 to 1920)
 AUTHORS Saeki K.
 TITLE Direct Submission
 JOURNAL Submitted (20-JUL-2000) Katsuhisa Saeki, Kao corporation,
 Biological Science Laboratory, Ichikamachi Akabane 2606, Haga,
 Tochigi 321-3497, Japan (E-mail: 387185@kagata-net.kao.co.jp,
 Tel: 81-285-68-7400, Fax: 81-285-68-7403)
 COMMENT On May 9, 2002 this sequence version replaced gi:12381938.
 FEATURES
 source
 1..1920
 /organism="Bacillus sp. 9860"
 /mol_type="genomic DNA"
 /strain="9860"
 /db_xref="taxon:133778"
 1..1920
 /gene="PROA"
 1..1920
 /gene="PROA"
 /codon_start=1
 /transl_table=11
 /product="protease"
 /protein_id="BAB21266.2"
 /db_xref="GI:20521153"
 /translations="MRKKVFLSVLSAAALSTVALNPSAGDARTFDLDFKGIQTTT
 DVSGFQRTQGTAAAFVSENVKLLKLETPANNKLIHVQNGPILFETKQK
 LETTGKILYIPDYAVIEYEDGVQSVRSIEHVESVPELPKYKIDPOLFTKCAST
 LKVALADTKONKKEVQLGIEIEIAQYASNDVHVIATKPEYKMNVDVARGIVKADVA
 OSSGLIGQSQIIVAVDTGLDTSNDSSEHAEFRGKILALYALGRTNNANDNGHGH
 VAGSVLNGAINTKGMAPQANLVFQSIWDSGGIGLGPLNLQILFQAPFAGARIHGTH
 WGAANVAGTTDSRNVDDYVRKNDMTILFAAGNERENGTTISAPGTAKNAITVGTATN
 LRPFSGYADININHVAFSSRGPTKDGRIKPDVMAFGTYILSARSLAPDSFVANH
 SKYAMGTSMTATVAGNVAQLREHFVKNRGTTPKPSLLKALIAAGADVGLGYPNG
 NQGMRTVTLKSLNVAYNVESSALSTSQKATFTTATAGKPLKISLVNSDAPASTAS
 VTLVNDLVLITAPNGTRYGVNDFSAFFDNNDGRNVNENVFINSQSGTYTIEVQAY
 NVPVGPQNFSIAVN"

ORIGIN
 Alignment Scores:
 Pred. No.: 1.05e-135 Length: 1920
 Score: 2237.00 Matches: 433
 Percent Similarity: 99.77% Conservative: 0
 Best Local Similarity: 99.77% Mismatches: 1
 Query Match: 99.64% Indels: 0
 DB: 1 Gaps: 0

US-09-985-689A-2 (1-434) x AB046403 (1-1920)

QY 1 AsnAspValAlaArgGlyIleValIleAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
 Db 616 AATGATGGCCAGAGGATTGTCAAAGCGGATGTGGCAGACAGCAGTACCGTTGTAT 675
 QY 21 GlyGlnGlyGlnIleValAlaValAlaAspThrClyLeuAspThrGlyArgAsnAspSer 40
 Db 676 GCACAAAGCCAGATGTTCGCGAGTTGCCGACTGGATTGGATACAGAAACGACAGT 735

QY 41 SerMetHisGluAlaPheArgGlyIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
 Db 736 TCGATGCTAGACCTTCGGGTAAATAAACACACTATATGCACTGGTTCGGACGAAAT 795
 QY 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
 Db 796 AATGCGAATGATACGAACGCTCATGTGATCCCATGTGGCAGGTTCGGTATTAGAAATGCG 855
 QY 81 AlaThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
 Db 856 GCACAGCAATTAAGAAATGGCACCTCAAGCGAATTCGGTTTTCAATCCATCATGATGATGC 915
 QY 101 SerGlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaPheSer 120
 Db 916 AGTGGTGGGCTTGGAGGCTTGCCTTCAATCTGCAACCTTATTCAGCAGCATTCAGT 975
 QY 121 AlaGlyAlaArgIleHisThrAsnSerTyrGlyAlaAlaValAsnGlyValAlaTyrThr 140
 Db 976 GCAGTGGCCAGATTTCATCAAACTCTCGGGGGGAGCGGTGAATGGGGCTTACACGACA 1035
 QY 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
 Db 1036 GATTCCAGAAATGTGATGACTATGTAAAGAAAAATGATATGACGATTCCTTTTCGCGCT 1095
 QY 161 GlyAsnGlyGlyProAsnGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
 Db 1096 GSGAATGAAGGCGAAGCGGGTACCATCATGTCACCTGGTACCGCTAAACGCCATA 1155
 QY 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
 Db 1156 ACAGTGGCGCAACCGAAAAACCTGGCTCCAAAGCTTCGGTTCCTATGCGATAAATTAAC 1215
 QY 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
 Db 1216 CACGTTGCACAGTTCTCTTCCTGGCCCGACAAAGATGGCGAATCAGCGCTGATGTC 1275
 QY 221 MetAlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
 Db 1276 ATGGCGCCAGGACATACATATTTATCAGCAAGATCTTCTCTTGCACCCGATTCCTCCTTC 1335
 QY 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIle 260
 Db 1336 TGGCGCAATCATCAGCAAAATATGCTATATGGGTGGACGCTCCATGCGCAACCGCAT 1395
 QY 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
 Db 1396 GTTGGCGGGAATGTTCACAGCTCCGTGAGCATTTTGTGAAAAATAGAGGAATCACTCCT 1455
 QY 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspValGlyLeuGlyTyr 300
 Db 1456 AAGCCTTCCTATTGAAACGAGCTTTGATTGTCAGGTGCTGCTGATGTTGGATTGGTTAT 1515
 QY 301 ProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
 Db 1516 CCGAACCGAAACCAAGGATGGGCGGAGTGACCCCTGGATAAATCGTTGAACGTTGCCATT 1575
 QY 321 ValAsnGlySerSerAlaLeuSerThrSerGlnLysAlaThrTyrThrPheThrAlaThr 340
 Db 1576 GTGAACGATTCAGTCCCTCATCACTAGCAAAAAAGCGACATATACCTTTACTGCAACG 1635
 QY 341 AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrAla 360
 Db 1636 GCGGCAAGCCATGTGAAATCTCCCTGGTATCGGTGGATGCCCTCGCAAGCAGCTACTCCT 1695
 QY 361 SerValThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrArgTyr 380
 Db 1696 TCTGTAACTCTGGTCAATGATTTGGATTGGTTCATTACAGACCAACCAACGAAGATAT 1755
 QY 381 ValGlyAsnAspPheSerAlaProPheAspAsnAsnTrpAspGlyArgAsnValGlu 400
 Db 1756 GTCGGGAATGACTTCTCAGCACCATTTGACATAAATCGGATGGCGCGCAATAACAGTAA 1815
 QY 401 AsnValPheIleAsnSerProGlnSerGlyThrTyrThrIleGluValIleAlaTyrAsn 420

Db 1816 AATGATTTATTAATTCGCCCAAGTGGACATATACCATTTGAGTGGACGATATAAT 1875
Qy 421 ValProValGlyProGlnAsnPheserLeuAlaIleValAsn 434
Db 1876 GTGCCGGTGGACACCAAACTTCTCGTTGGCAATTTGTGAAC 1917
RESULT 2
AR368116
LOCUS AR368116 1920 bp DNA linear PAT 12-SEP-2003
DEFINITION Sequence 3 from patent US 6376227.
ACCESSION AR368116
VERSION AR368116.1 GI:34601777
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1920)
AUTHORS Takaiwa, M., Okuda, M., Saeki, K., Kubota, H., Hitomi, J., Kageyama, Y.,
Shikata, S. and Nomura, M.
TITLE Alkaline protease
JOURNAL Patent: US 6376227-A 3 23-APR-2002;
FEATURES Location/Qualifiers
source 1..1920
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN
Alignment Scores: 1,05e-135 Length: 1920
Pred. No.: 2327.00 Matches: 433
Score: 2327.00
Percent Similarity: 99.77% Conservatives: 0
Best Local Similarity: 99.77% Mismatches: 1
Query Match: 99.64% Indels: 0
DB: 6 Gaps: 0
US-09-985-689A-2 (1-434) x AR368116 (1-1920)
Qy 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnSerSerTyrClyLeuTyr 20
Db 616 AATGATGTGCCAGAGATTTGTCAAAGCGGATGTGGCACAGACAGCTACCGTTTGTAT 675
Qy 21 GlyGlnGlyClnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 676 GGAACAAGCCAGATTGTCCAGTGTCCGATATCGATTGGATACAGGAAGAACACGAGT 735
Qy 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 736 TCGATGCAATGAAGCCATTCGCGGTAAATAACAGCACTATATGCACTGGGTGCGACGAAT 795
Qy 61 AsnAlaAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
Db 796 AATGCCAATGATACGAACCGTATGTGTACCATGTGGCAGGTTCCGTTATAGGAATGGC 855
Qy 81 AlaThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheClnSerIleMetAspSer 100
Db 856 GCAACGAATAAAGAAATGGCACTCAAGCGAATCTGGTTTTCATCCATCATGTGATAGC 915
Qy 101 SerGlyLeuGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaPheSer 120
Db 916 AGTGGTGGGCTGGAGCTTGCCTTCAATCTCAAACTTATTCAGCCAGCAATTCAGT 975
Qy 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyValAlaValAsnGlyAlaTyrThrThr 140
Db 976 GCAGGTGCCAGATTTCATAAACTCTGGGGGCGAGCGGTGAATGGGCTACACGACA 1035
Qy 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
Db 1036 GAITCCAGAAATGTGGATGACTATGAAGGAAAAATGATATGACGATCTTTTCGCGGCT 1095
Qy 161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
Db 1096 GGGNATGAAGCCGACCGCGGTACCATCAGTGCACCTGGTACCGCTAAACACGCCATA 1155

Qy 181 ThrValGlyValAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
Db 1156 ACAGTCGGCGCAACCGAAACCTCGCTCAAGCTTCGGTTCCTATGACGATATAATTAAC 1215
Qy 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
Db 1216 CACGTTGCACAGTCTCTTCCGTGGCCGCAAAAGATGGCGGAATCAAGGCTCATGTC 1275
Qy 221 MetAlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
Db 1276 ATGGCCCGCAGGACATACATTTTATCAGCAAGATCTTCTTTCACCCGATTCCTCTTC 1335
Qy 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260
Db 1336 TGGCGCAATCATGACAGCAATATGCCTATATGGGTGGAACTCCATGTCACACACCGATT 1395
Qy 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
Db 1396 GTTGGCGGGAATGTTGCACAGCTCCGTGAGCAATTTGTGAAAAATAGAGGAATCACTCT 1455
Qy 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspValGlyLeuGlyTyr 300
Db 1456 AAGCCTTCCCTATTGAAGCAGCTTTGATTGACAGGTGCTGCTGATGTTGGATTGGTTAT 1515
Qy 301 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
Db 1516 CCGAACCGAAACAGGATGGGCGCGAGTGACCTCGATAAATCGTTGAACGTTGCTAT 1575
Qy 321 ValAsnGluSerSerAlaLeuSerThrSerGlnLysAlaThrTyrThrPheThrAlaThr 340
Db 1576 GTGAACGAATCCAGTGCCTTATCAACTAGCAAAAGACGACATATACCTTTACTGCAACG 1635
Qy 341 AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrAla 360
Db 1636 GCGGCGAAGCCATTGAAATCTCCCTGATGTGCGATGCCCTCCGATGCGCAAGCACTACTGCT 1695
Qy 361 SerValThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrArgTyr 380
Db 1696 TCTGTAAACCTGTGTCATGATTGGATTGGTTCATTACAGCAACCAACGAAACAGATAT 1755
Qy 381 ValGlyAsnAspPheSerAlaProPheAspAsnAsnTrpAspGlyArgAsnValGlu 400
Db 1756 GTCCGGGAATGACTTCTCAGCACCATTTGACAATACTGGGATGGCCGCAATAACGATAGAA 1815
Qy 401 AsnValPheIleAsnSerProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
Db 1816 AATGATTTATTAATTCGCCCAAGTGGAAACATATACCATTCAGGTGCAAGCATATAAT 1875
Qy 421 ValProValGlyProGlnAsnPheserLeuAlaIleValAsn 434
Db 1876 GTGCCGGTGGACCAAAACCTTCTCGTTGGCAATTTGTGAAC 1917
RESULT 3
AR368118
LOCUS AR368118 1923 bp DNA linear PAT 12-SEP-2003
DEFINITION Sequence 7 from patent US 6376227.
ACCESSION AR368118
VERSION AR368118.1 GI:34601779
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1923)
AUTHORS Takaiwa, M., Okuda, M., Saeki, K., Kubota, H., Hitomi, J., Kageyama, Y.,
Shikata, S. and Nomura, M.
TITLE Alkaline protease
JOURNAL Patent: US 6376227-A 7 23-APR-2002;
FEATURES Location/Qualifiers
source 1..1923
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN

| | | | |
|--|--------|---|------------------|
| Alignment Scores: | | 4.07e-133 | Length: 1923 |
| Pred. No.: | Score: | 2197.00 | Matches: 420 |
| Percent Similarity: | | 99.77% | Conservative: 13 |
| Best Local Similarity: | | 96.77% | Mismatches: 1 |
| Query Match: | | 97.86% | Indels: 0 |
| DB: | | 6 | Gaps: 0 |
| US-09-985-689A-2 (1-434) x AR368118 (1-1923) | | | |
| QY | 1 | AsnAspValAlaAargGlyLeValLysAlaAspValAlaGlnSerSerThrGlyLeuTyr 20 | |
| DB | 619 | AATGATGTTGGCGTGGAAATGTCAAAGCGGATGTGGCTCAGACGAGCTACGGGTGTAT 678 | |
| QY | 21 | GlyGlnGlyGlnLeValAlaValAlaAspThrGlyLeuAspThrGlyValArgAspSer 40 | |
| DB | 679 | GGCAAGGACAGATCGTAGGGTTGCCGATACAGGCTTGATACAGGTCCCATGACAGT 738 | |
| QY | 41 | SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60 | |
| DB | 739 | TCATGTCATGAACCTTCGCGGAAAAATTACTGCATTATATGTCATTGGGACGCAAT 798 | |
| QY | 61 | AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80 | |
| DB | 799 | AATGCCAATGATACGAATGCTCATGTGTACGATGCTGGCTGGCTCCGTATTAGAAACGGC 858 | |
| QY | 81 | AlaThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100 | |
| DB | 859 | TCACATTAATAAGGAATGGCGCTCAGCGAACTAGTCTTCCAACTATCATGTGATAGC 918 | |
| QY | 101 | SerGlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaPheSer 120 | |
| DB | 919 | GGTGGGGACTTGGAGGACTACCTCGAATCTGCAACCTTATTCAGCCAGCATACAGT 978 | |
| QY | 121 | AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyrThrThr 140 | |
| DB | 979 | GCTGGTCCAGAACTATACAACTCTGGGAGCAGCAGTGAATGGGGCTTACACAACA 1038 | |
| QY | 141 | AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160 | |
| DB | 1039 | GATTCAGAAATGTGGATGACTATGTGCGCAAAATGATATGACGATCCTTTTCGGTGCC 1098 | |
| QY | 161 | GlyAsnGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180 | |
| DB | 1099 | GGGAATGAAGGACCGAACCGCGGAACCACTAGTCCACAGGCAGCTAATAATGCAATA 1158 | |
| QY | 181 | ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200 | |
| DB | 1159 | ACAGTCGGAGCTACGGAAACCTCCGCCAAGCTTTGGGTCTTATGGGCAATATCAAC 1218 | |
| QY | 201 | HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220 | |
| DB | 1219 | CATGTGGCAGATCTCTTCACGTGGACCGAAGGATGACGATCAACCGGATGTC 1278 | |
| QY | 221 | MetAlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240 | |
| DB | 1279 | ATGGCACCGGGAACGTTCTATCATATCAGCAAGATCTTCTCTTGCACCGGATTCCTCCTC 1338 | |
| QY | 241 | TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProile 260 | |
| DB | 1339 | TGGGCGAACCATGACAGTAATAATATCATACATGGGTGGAACGTCATCGCTACACCGATC 1398 | |
| QY | 261 | ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280 | |
| DB | 1399 | GTTCGTGGAAACGTGGCAGCTTCGTGAGCATTTTGTGAAAAACAGAGGCATCACACCA 1458 | |
| QY | 281 | LysProSerLeuLeuLysAlaAlaLeuLeuAlaGlyAlaAlaAspValGlyLeuGlyTyr 300 | |
| DB | 1459 | AAGCTCTCTATTAAAAAGCGCACTGATTCGGGTGAGTGATCGGCTTGGCTTAC 1518 | |
| QY | 301 | ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320 | |

| | | | |
|--|------|---|--|
| DB | 1519 | CCGAACGGTAACCAAGGATGGGACGAGTGACATTGGATAAATCCCTGAACGTTGCCTAT 1578 | |
| QY | 321 | ValAsnGluSerSerAlaLeuSerThrSerGlnLysAlaAlaThrTyrThrPheThrAlaThr 340 | |
| DB | 1579 | GTGAACGATCCAGTCTCTATCCACGCAAAAGCGAGCTACTCGTGTACTGTACT 1638 | |
| QY | 341 | AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrThrAla 360 | |
| DB | 1639 | GCGGCAAGCCTTTGAAATCTCCCTGGTATGCTGTGATGCCCTCGGAGCACCACTGCT 1698 | |
| QY | 361 | SerValThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrArgTyr 380 | |
| DB | 1699 | TCGTACAGCTTGTCAATGATCTGGACCTTGTCAATCCGTCCTCAATGGCACACAGTAT 1758 | |
| QY | 381 | ValGlyAsnAspPheSerAlaProPheAspAsnAsnTrpAspGlyArgAsnValGlu 400 | |
| DB | 1759 | GTAGAAATGACATTACTTCGCATACATGATACTGGGATGCGCGCAATACGTAAGA 1818 | |
| QY | 401 | AsnValPheIleAsnSerProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420 | |
| DB | 1819 | AATGATTTTATTAATGACCAACGCGGACGTATACAATTGAAGTACAGGCTTATAAC 1878 | |
| QY | 421 | ValProValGlyProGlnAsnPheSerLeuAlaIleValAsn 434 | |
| DB | 1879 | GTACCGGTTGGACACAGAACTTCTCGTTGGCAATTGTGAAT 1920 | |
| RESULT 4 | | | |
| AB084155 1923 bp DNA linear BCT 02-SEP-2003 | | | |
| LOCUS Bacillus sp. KSM-9865 gene for protease, complete cds. | | | |
| DEFINITION AB084155 | | | |
| ACCESSION AB084155.1 GI:34392386 | | | |
| VERSION | | | |
| KEYWORDS | | | |
| SOURCE | | | |
| ORGANISM | | | |
| REFERENCE | | | |
| AUTHORS | | | |
| TITLE | | | |
| JOURNAL | | | |
| REFERENCE | | | |
| AUTHORS | | | |
| TITLE | | | |
| JOURNAL | | | |
| FEATURES | | | |
| source | | | |
| CDS | | | |
| ORIGIN | | | |

| | | | |
|--|--|--|-----------------|
| Alignment Scores: | | 4.73e-133 | 1923 |
| Pred. No.: | 2196.00 | Matches: | 420 |
| Score: | 99.54% | Conservative: | 12 |
| Best Local Similarity: | 96.77% | Mismatches: | 2 |
| Query Match: | 97.82% | Indels: | 0 |
| DB: | 1 | Gaps: | 0 |
| US-09-985-689A-2 (1-434) x AB084155 (1-1923) | | | |
| QY | 1 | AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnSerSerTyGlyLeuTyr | 20 |
| Db | 619 | AATGATGTCACGTGCAATGTCAGGCGGATGTCGCGCAGACGAGCTACGGGTTGTAT | 678 |
| QY | 21 | GlyGlnGlyGlnIleValAlaAlaAspThrGlyLeuAspThrGlyArgAsnAspSer | 40 |
| Db | 679 | GGCAAGGACAGATCGTAGCGGTTGCCGATACAGGCGCTTATACAGGTGCAATGACAGT | 738 |
| QY | 41 | SerMetHisGluAlaPheAspGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn | 60 |
| Db | 739 | TCGATGCGATGAGCTTCCGGGGAAATTAATCTGCTATATATGATTTGGGACGAGAT | 798 |
| QY | 61 | AsnAlaAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly | 80 |
| Db | 799 | AATGCCAATATACCAATGTCATGTTACGATGTGGCTGGCTCCGTTATTAGGAAACGGC | 858 |
| QY | 81 | AlaThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer | 100 |
| Db | 859 | TCCTACTAATAGGAATGGCGCTCAGCGCAATCTAGCTTCCAACTATCATGGATAGC | 918 |
| QY | 101 | SerGlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaPheSer | 120 |
| Db | 919 | GGTGGGGGACTTGGAGGACTACCTTCGAATCTGCAAACTTATTTCAGCCAGCATACAGT | 978 |
| QY | 121 | AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyrThr | 140 |
| Db | 979 | GCTGGTGCAGAAATTCATACAAATCTCTGGGGAGCAGCAGTGAATGGGCTTACAAACA | 1038 |
| QY | 141 | AspSerArgAsnValAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla | 160 |
| Db | 1039 | GATTCAGAAATGTGGATGACTATGTGGCAAAATGATATGACGATCCCTTTTCGTGCC | 1098 |
| QY | 161 | GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle | 180 |
| Db | 1099 | GGGAATGAAGACCGAAGCGCGCAACCATCATGTCACCGACGACAGCTAAATAATGCAATA | 1158 |
| QY | 181 | ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn | 200 |
| Db | 1159 | ACAGTCGGAGTACGGAAACCTCGCCCAAGCTTCGGTCTTATCGGACAAATATCAAC | 1218 |
| QY | 201 | HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal | 220 |
| Db | 1219 | CATGTGGCACAGTTCTCTTTCACGTGGACCGCAAGGATGGACGGATCAACCGGATGTC | 1278 |
| QY | 221 | MetAlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe | 240 |
| Db | 1279 | ATGGCACCGGACGCTTCATCTACTATCAGCAGATCTTCTTTGACCGGATTCCTCTTC | 1338 |
| QY | 241 | TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle | 260 |
| Db | 1339 | TGGGCGAACCATGACAGTAAATATGATACATGATGGTGGAACTGCTCCATCGCTTACACCGATC | 1398 |
| QY | 261 | ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro | 280 |
| Db | 1399 | GTGTGTGGAACGTGGCAGCTTCGTGAGCATTTTGTGAAATAACAGAGGATCACACCA | 1458 |
| QY | 281 | LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspValGlyLeuGlyTyr | 300 |
| Db | 1459 | AAGCCTTCTCTATTAAAGCGCACTGATTGCGGTGCGAGCAGACATCGGCTTTGGCTAC | 1518 |
| QY | 301 | ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr | 320 |
| Db | 1519 | CCGAACGGTAACCAAGGATGGGACAGTGGACATTTGGATAAATCCCTAAACGTTGCCCTAT | 1578 |
| QY | 321 | ValAsnGluSerSerAlaLeuSerThrSerGlnLysAlaThrTyrThrPheThrAlaThr | 340 |
| Db | 1579 | GTGAACGAGTCCAGTTCTCTATCCACGAGCAAAAGCGAGTACTCTGTTTACTGCTACT | 1638 |
| QY | 341 | AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrAla | 360 |
| Db | 1639 | GCCGGCAAGCCTTTGAAGATCTCCTGTGTATGCTGATGCCCTGCGAGCAACTGCT | 1698 |
| QY | 361 | SerValThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrArgTyr | 380 |
| Db | 1699 | TCCGTAAACGCTTGTCAATGACCTGCGACCTTGTCTATTACCGCTCCAAATGGCACAATAT | 1758 |
| QY | 381 | ValGlyAsnAspPheSerAlaProPheAspAsnAsnTrpAspGlyArgAsnValGlu | 400 |
| Db | 1759 | GTTGGAATGACTTACTTCGCCATACATAATACTACCTGGGATGCCCGCAATAACGTAGAA | 1818 |
| QY | 401 | AsnValPheIleAsnSerProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn | 420 |
| Db | 1819 | AATGTATTATTAAATGCGCCCAAGCGGACGATATACAAATTGAGGTACAGGCTTATAAC | 1878 |
| QY | 421 | ValProValGlyProGlnAsnPheSerLeuAlaIleValAsn | 434 |
| Db | 1879 | GTACCGGTTGGACCAACACACCTTCTCGTTGCCAATTGTGAAC | 1920 |
| RESULT 5 | | | |
| AX839476 | 1305 bp | DNA | linear |
| AX839476 | Sequence 2 from Patent EP1347044. | | PAT 15-DEC-2003 |
| DEFINITION | AX839476 | | |
| ACCESSION | AX839476.1 | GI:39922766 | |
| VERSION | AX839476.1 | GI:39922766 | |
| KEYWORDS | Bacillus sp. KSM-KP43 | | |
| SOURCE | Bacillus sp. KSM-KP43 | | |
| ORGANISM | Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus. | | |
| REFERENCE | 1 | | |
| AUTHORS | Okuda, M.K., Saito, T.K., Saito, K.K., Sumitomo, N.K., Izawa, Y.K., Saeiki, K.K., Kobayashi, T.K. and Nomura, M.K. | | |
| TITLE | Alkaline protease | | |
| JOURNAL | Patent: EP 1347044-A 2 24-SEP-2003; | | |
| FEATURES | Kao Corporation (JP) | | |
| source | Location/Qualifiers | | |
| CDS | 1..1305 | | |
| | /organism="Bacillus sp. KSM-KP43" | | |
| | /mol_type="unassigned DNA" | | |
| | /db_xref="taxon:109322" | | |
| | <1..1305 | | |
| | /note="unnamed protein product" | | |
| | /codon_start=1 | | |
| | /transl_table=11 | | |
| | /protein_id="CAE85146.1" | | |
| | /db_xref="GI:39922767" | | |
| | APRKITALIYALGRTNNANDTNGHGHVAGSVLNGSTNGKMAPOANLVFOSIMDSGG | | |
| | GIGGLPSNLQTLFSQAYAGARIHTNSWGAAVNGAYTTDSRNVDDYVRKNDMTILFAA | | |
| | GNENPGNGTISAPGTAKNAITVGTENLRPSFYADINHVAPQSSGPTKDGRIKP | | |
| | DYVAPGFTLSASPSLIPDSFWANHDSKYAVMGTSMTATVAGNVQALREHFVYKR | | |
| | GITPKPSLLKALIAAADIGLGYPNGQGWGRVTLDKSLNAVYVNESLSLSQKAT | | |
| | YFTATACKPLKISLVNSDAPASTASVTLVNDLVLITAPNGTQYVGNDFTSFYNDN | | |
| | WDGRNVNFINAFQSGYITIEVQAVNPVGPQFSLAYN" | | |
| ORIGIN | | | |
| Alignment Scores: | 6.17e-133 | Length: | 1305 |
| Pred. No.: | 2191.00 | Matches: | 419 |
| Score: | 99.54% | Conservative: | 13 |
| Best Local Similarity: | 96.54% | Mismatches: | 2 |
| Query Match: | 97.59% | Indels: | 0 |
| DB: | 6 | Gaps: | 0 |
| US-09-985-689A-2 (1-434) x AX839476 (1-1305) | | | |
| QY | 1 | AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnSerSerTyGlyLeuTyr | 20 |

1 AATGATGTTGGCGTGAATGCTCAAGCGGATGTTGGCTCAGACAGCTACGGGTTCTAT 60
 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAspSer 40
 61 GGACAGGACAGATCGTAGCGGTTCGCCATACAGGCTTGTATACAGTTCGCAATGACAGT 120
 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyAlaLeuGlyArgThrAsn 60
 121 TCGATGATGAAGCCTTCGCGGGGAAATTAATGCTATATATGCTATGATGCGGACGCAAT 180
 61 AsnAlaAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
 181 AATGCCAATGATACCAATGGTTCATGCTACGATGCTGGCTCGCTTCGTTATGAGAAACGC 240
 81 AlaThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
 241 TCCACTAATAAGGAATGGCGCTCAGCGCAATCTAGTCTTCCAATCTATCATGATGATG 300
 101 SerGlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaPheSer 120
 301 GTGGGGGACTTGGAGGACTCTCTCGAATCTGCAACCTTATTGAGCAAGCATACAGT 360
 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyValaTyThrThr 140
 361 GTGTGCTCCAGAACTCATACAACTCTCTGGGAGCAGCAGTGAATGGGCTTACACAACA 420
 141 AspSerArgAsnValAspAspTyValArgLysAsnAspMetThrIleLeuPheAlaAla 160
 421 GATTCAGAAATGTGATGACTATGTGCGCAAAATGATATGACCATCTTTTCCTGCC 480
 161 GlyAsnGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
 481 GGAATGTAGAGCAGCGGACCGGACCACTCAGTGCCACGAGCAGCAGCTAAATGCAATA 540
 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyAlaAspAsnIleAsn 200
 541 ACAGTCGAGACTACGGAAACCTCCGCCCAAGCTTTGGGTCTTATGCGGACAAATCAAC 600
 201 HisValAlaGlnPheSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
 601 CATGTGGCAGATTCCTTCATGTGGACCCGACCAAGATGGAGGATCAACCGGATGTC 660
 221 MetAlaProGlyThrTyIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
 661 ATGGACCGGGAACGTTTCATCTATCATCAGCAAGATCTCTCTTGACCGGATTCCTCTTC 720
 241 TrpAlaAsnHisAspSerLysTyAlaTyMetGlyThrSerMetAlaThrProIle 260
 721 TGGCGGAACCATGACAGTAATATGCATATCATGGTGGACGTCCTCATGCTACACCGATC 780
 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
 781 GTTCCTGGAAACGTGGCACAGCTTCGTGAGCAATTTGTGAAAAACAGAGGCATCACACCA 840
 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspValGlyLeuGlyTy 300
 841 AAGCCTTCTCTATTAAAGCGGCACTCAATTCGCGTGCATGACCTGACATGCGCCTTGCTAC 900
 301 ProAsnGlyAsnGlnGlyTyTrpGlyArgValThrLeuAspLysSerLeuAsnValaTy 320
 901 CCGAACGGTAACCAAGGATGGGACGAGTGACATTTGATAAATCCCTGAACTGCGCTAT 960
 321 ValAsnGluSerSerAlaLeuSerThrSerGlnLysAlaThrTyThrPheThrAlaThr 340
 961 GTGAACGAGTCCAGTCTCTATCCACCGCAAAAGCGAGTACTCGTTTACTGCTACT 1020
 341 AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrAla 360
 1021 CCGCGCAAGCCTTTGAAATCTCCTCTGTAAGTCTGATGCCCTCGGAGCAACACTGCT 1080
 361 SerValThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrArgTy 380

1081 TCCTAAACGCTTGTCAATGATCTGGACCTTGTCTATCCGCTCCAAATGGCACAGTAT 1140
 381 ValGlyAsnAspPheSerAlaProPheAspAsnAsnTrpAspGlyArgAsnAsnValGlu 400
 1141 GTAGGAATGACTTTACTTCGCCATACATGATAACTGGGATGGCGCAATTAACGTAGAA 1200
 401 AsnValPheIleAsnSerProGlnSerGlyThrTyThrIleGluValGlnAlaTyAsn 420
 1201 AATGATTTATTATATGCACCAACAGCGGACGTATACAAATGAGGTACAGCTTATAAC 1260
 421 ValProValGlyProGlnAsnPheSerLeuAlaIleValAsn 434
 1261 GTACCGGTTGGACCAAGACCTTCTCGTTGGCAATTTGTAAT 1302
 RESULT 6
 AB051423
 LOCUS
 DEFINITION
 AB051423
 ACCESSION
 VERSION
 AB051423.2 GI:20521154
 KEYWORDS
 SOURCE
 ORGANISM
 Bacillus sp. KSM-KP43
 Bacillus sp. KSM-KP43
 Bacteria; Firmicutes; Bacilliales; Bacillaceae; Bacillus.
 REFERENCE
 1 Itoh, S. and Saeki, K.
 new protease
 TITLE
 JOURNAL
 2 (bases 1 to 1923)
 AUTHORS
 Saeki, K.
 DIRECT SUBMISSION
 Submitted (21-NOV-2000) Katsuhisa Saeki, KAO CORPORATION;
 2606, AKABANE, ICHIKAINACHI, HAGA, TOCHIGI 321-3486, Japan
 (E-mail: 387185@katsanet.kao.co.jp, Tel: 81285687471 (ex. 7471),
 Fax: 81285687403)
 On May 9, 2002 this sequence version replaced gi:14164344.
 COMMENT
 FEATURES
 source
 Location/Qualifiers
 1..1923
 /organism="Bacillus sp. KSM-KP43"
 /mol_type="genomic DNA"
 /strain="KP43"
 /db_xref="taxon:109322"
 1..1923
 /gene="PROF"
 1..1923
 /gene="PROF"
 /codon_start=1
 /transl_table=11
 /product="protease"
 /protein_id="BAB55674.2"
 /db_xref="GI:20521155"
 /translation="MRKKKVFSLVLSAAAILSTVALSNPSAGARNFDLDFKGIQTT
 TDAGFSKQGTGAALFVSENVKLPKGLKLETPANNKLHIQFNGPILBETKQ
 QLEKTGAKIDYIFDYAYIVEYEDVKSATSTIEHVSVPYPIVYRDPQLFTKAS
 ELKAVALDTQKNKEVQLRGEIQIALSNVLYITAKPEYKVMNDVARGIVKADV
 AQSGYGLGGQIVAVADTGLDTRGNSDWEAFRGKITALIALGRNNDNTHGHT
 HVAQSVLNGSTNKGMAQANLVQISMDSGGLGGLPSNLQTLFQSAISAGARIHTN
 SWGAANVAGTYTDSRNVDDYVRKNDMTILFAAGNEGNGGTISAPGTAKNAITVGAH
 NLRPSFGYADNINHVAFSGSRGPTKGRIPKPDVMAPTTILSARSLAPDSSFWANH
 DSKAYNGGTSMATPIVAGNVAQLREHFVKNRGTPKPSLLKALIAAGADIGLYEN
 GQGWGRVLLKSLNVAYNHSSLSSTSQKATYSFTATAGKPKLSLWSDAPASTTA
 SVTLVNDLDLVITAPNGTQYVGNDFTSFYNDNWDGRNVNFNAPQSCTTYTIEVQA
 YNVPVPGPTFSLAIVN"

ORIGIN

Alignment Scores:
 Pred. NO.: 9,956-133 Length: 1923
 Score: 2191.00 Matches: 419
 Percent Similarity: 99.54% Conservative: 13
 Best Local Similarity: 96.54% Mismatches: 2
 Query Match: 97.59% Indels: 0
 DB: 1 Gaps: 0

Db 979 GCTGGTCCAGAAATTCATAAACTCTCGGGAGCAGCAGTGAATGGGCTTACACAACA 1038

Qy 141 AspSerArgAsnValAspAspTyrValArgLyAsnAspMetThrIleLeuPheAla 160

Db 1039 GATTCCAGAAATGTGGATGACTATGTGCGCAAAATGATATGACGATCTTTTCGCTGCC 1098

Qy 161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLyAsnAla 180

Db 1099 GGGAAATGAAGACCGAAGCGGGACCATCAGTCACCGACGACAGCTAAATGCAATA 1158

Qy 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200

Db 1159 ACATGCGAGCTACGAAACCTCCGCCAAGCTTTGGGTCTTATCGGACATATCAAC 1218

Qy 201 HisValAlaGlnPheSerArgGlyProThrIleAspGlyArgIleIysProAspVal 220

Db 1219 CATGTGGCACAGTTCTTTCACGTGGACCGACCAAGGATGGACGATCAACCGGATGTC 1278

Qy 221 MetAlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerPhe 240

Db 1279 ATGGCACCGGGAACGTCATACTATCAGCAAGATCTTCTCTCCACCGGATTCCTCCTC 1338

Qy 241 TrpAlaAsnHisAspSerIysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260

Db 1339 TGGCGCAACCATGACAGTAATATGCATATCATGGTGGACGTCCTGCTACACGATC 1398

Qy 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValIysAsnArgGlyIleThrPro 280

Db 1399 GTTGTGAAACGTTGGCACAGCTTCGTGAGCATTTTGTGAAACAGAGGATCAACCA 1458

Qy 281 LysProSerLeuLeuIysAlaAlaLeuIleAlaGlyAlaAlaAspValGlyLeuGlyTyr 300

Db 1459 AAGCCTCTCTATTAAAGCGCACTGATTCGCGGTGACGCTGACATCGCGCTTGGCTAC 1518

Qy 301 ProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspIysSerLeuAsnValAlaTyr 320

Db 1519 CCGAACCGTAACCAAGATGGGACGAGTGACATTCGATAAATCCCTGAACGTGCTAT 1578

Qy 321 ValAsnGluSerSerAlaLeuSerThrSerGlnIysAlaThrTyrThrPheThrAlaThr 340

Db 1579 GTGACGAGTCCAGTTCTCTATCCACCGCAAAAGCGACGACTCTGTTACTGCTACT 1638

Qy 341 AlaGlyIysProLeuIysIleSerLeuValTrpSerAspAlaProAlaSerThrAla 360

Db 1639 CGCGCAAGCCTTTGAAATCTCCCTGCTGATGCTGATGCCCTCGGACGACCACTGCT 1698

Qy 361 SerValThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrArgTyr 380

Db 1699 TCCGTACGCTGTCAATGATCTGGACCTTGTCTATCCGCTCCAAATGGCACACAGTAT 1758

Qy 381 ValGlyAsnAspPheSerAlaProPheAspAsnAsnTrpAspGlyArgAsnValGlu 400

Db 1759 GTAGGAATGACTTTACTTCGCCATACATGATACTGGATGGCGCAATAACGTAGAA 1818

Qy 401 AsnValPheIleAsnSerProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420

Db 1819 AATGTATTATTATGATCACCACCAAGCGGACGATACAAATGAGGTACAGGCTTATAAC 1878

Qy 421 ValProValGlyProGlnAsnPheSerLeuAlaIleValAsn 434

Db 1879 GTACCGGTTGGACACAGACCTTCTCGTTGGCAATGTGAAT 1920

RESULT 8

AB046406
 LOCUS AB046406 1302 bp DNA linear BCT 23-JAN-2001
 DEFINITION Bacillus sp. NV1 PROE gene for protease, partial cds.
 ACCESSION AB046406
 VERSION AB046406.1 GI:12381944
 KEYWORDS
 SOURCE Bacillus sp. NV1
 ORGANISM Bacillus sp. NV1
 Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

gene

CDS

ORIGIN

Alignment Scores:

Pred. No.:

Score:

Percent Similarity:

Best Local Similarity:

Query Match:

DS:

US-09-985-689A-2 (1-434) x AB046406 (1-1302)

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

1 (sites)
 Saeki, K., Okuda, M., Hatada, Y., Kobayashi, T., Ito, S., Takami, H. and
 Horikoshi, K.
 Novel oxidatively stable subtilisin-like serine proteases from
 alkaliphilic *Bacillus* spp.: enzymatic properties, sequences, and
 evolutionary relationships
 Biochem. Biophys. Res. Commun. 279 (2), 313-319 (2000)

20568675
 11118284
 2 (bases 1 to 1302)
 Saeki, K.
 Direct Submission
 Submitted (20-JUL-2000) Katsuhisa Saeki, Kao corporation,
 Biological Science Laboratory, Ichikaimachi Akabane 2606, Haga,
 Tochigi 321-3497, Japan (E-mail: 387185@kasanet.kao.co.jp,
 Tel: 81-285-68-7400, Fax: 81-285-68-7403)
 Location/Qualifiers
 1..1302
 /organism="Bacillus sp. NV1"
 /mol_type="genomic DNA"
 /strain="NV1"
 /db_xref="taxon:133781"
 1..1302
 /genes="PROE"
 <1..>1302
 /genes="PROE"
 /codon_start=1
 /transl_table=11
 /product="protease"
 /protein_id="BAB21269.1"
 /db_xref="GI:12381945"

/translation="NDVARGIVKADVAQSSVLYGQGVAVADTGLDTRNDSMH
 AFRGKIALYALGRNNDPNHGHTHVAGSVLNGTSKMGKMAQANLVFQSVMDNSG
 GGGPFSNVSTLFSQAYAGARITNSWGAFFVAYTDSRVDDYVRKNDMAVLFAA
 GNEGNGGTISAFKXNAITVGTENLRFSGYADNINVAOFSSRGPTKDGRIK
 DVMAPTFILSARSLPADPSFWANHDSKYAMGTSNATIVAGNVQALREHFIKPR
 GTPKPSLLKAAALIAAGTDIGLYPSGNQGRVTLDKSLNVAFNWTSLSLTOKAT
 YFTAQSGKPLAISLWSDAPASTASVTLVNDLDELVTAPNGIKYGVNDTFTPYDNN
 WGRNNVENVFVAPQSGTYTVEVQAVNPQGPQAFSLAIVN"

ORIGIN

Alignment Scores:

Pred. No.:

Score:

Percent Similarity:

Best Local Similarity:

Query Match:

DS:

US-09-985-689A-2 (1-434) x AB046406 (1-1302)

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Db 301 AATGCGGACTTGGCGAGTGCCTTCCAAATGTAAGTACATTATTACGCCGACATATAGT 360
Qy 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyrThr 140
Db 361 CCCGGTGCAGATCCATCAGACTCATGCGGAGCCCTGTAAACGAGGCGTACACTACT 420
Qy 141 AspSerArgAsnValAspAspTyrValArgIlyAsnAspMetThrIleLeuPheAlaAla 160
Db 421 GATTCAGAAACGCTAGACGATTATGTCGTAAAAATGATATGGCGGTTCTTTTTCGACG 480
Qy 161 GlyAsnGluGlyProAsnGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
Db 481 GGTACGAAAGGCGCGAATGGCGGCAATCAGTCTCTGTACCGGAAGATGCTATC 540
Qy 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
Db 541 ACAGTAGGGCAACAGAAACCTGGCCCAAGCTTTGGATCTTATGCTTGACACATCAAT 600
Qy 201 HisValAlaGlnPheSerSerArgGlyProThrIlyAspGlyArgIleLysProAspVal 220
Db 601 CATGTAGCACAGTTTTCTCCCGCGAGCTTCAAGAGGATGGACGTATCAACCCGGAGTA 660
Qy 221 MetAlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
Db 661 ATGGCACCAGGAACATTTATTTATTCGCAAGATCTTCTTTGGCTCCGGACTCTCATTC 720
Qy 241 TrpAlaAsnHisAspSerIlyThrAlaTyrMetGlyThrSerMetAlaThrProIle 260
Db 721 TGGGCAAAACCATGACAGCAATATCTTATATGGGTGGAAATCCATCGCCACACCGATT 780
Qy 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
Db 781 GTAGTGGTAACGTTGCAGATTACGTGAACATTTTCACAAAACAGAGGAATCACTCCT 840
Qy 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspValGlyLeuGlyTyr 300
Db 841 AAACCATCTCTGCTGAAGACAGCTCTTATTCGCGAGCACTGATATCGCTCTGGCTAT 900
Qy 301 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
Db 901 CCGAGTGGAAACCAAGGATGGGAGAGATACATTTGGACAGTCACTTAATGTAGCTTC 960
Qy 321 ValAsnGluSerSerAlaLeuSerThrSerGlnLysAlaThrTyrThrPheThrAlaThr 340
Db 961 GTAAATGAACAAAGCTCTTTACTACTAACCAAAAGGCTAGCTATTCATTTTACTGCACA 1020
Qy 341 AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrAla 360
Db 1021 TCAGGCAACCTTTGAAGATTTCATTTGGTTGGTCTGATGCACCGCAAGTACTTCGCA 1080
Qy 361 SerValThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrArgTyr 380
Db 1081 TCGGTTACATTGGTGAATGATCTGGATCTGGTGATTACAGCTCCAAATGGAAACAGTAT 1140
Qy 381 ValGlyAsnAspPheSerAlaProPheAspAsnAsnTrpAspGlyArgAsnValGlu 400
Db 1141 GTTGAACAGCTTTACTCTCTCCCTATGATAATACTGGATGGACGTAAACAATGTAGAG 1200
Qy 401 AsnValPheIleAsnSerProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
Db 1201 AACGTGTTTATCAATGCTCCGCAAGCGAAGCTATACAGTTGAGGTACAGGCTTACAAT 1260
Qy 421 ValProValGlyProGlnAsnPheSerLeuAlaIleValAsn 434
Db 1261 GTTCCACAAGGGCCGAGCGCTTTCTTTGGCTATTGTGAAC 1302

RESULT 9

AR069954
LOCUS 3003 bp DNA linear PAT 18-FEB-2000
DEFINITION Sequence 41 from patent US 5891701.
ACCESSION AR069954
VERSION AR069954.1 GI:7220842
KEYWORDS

SOURCE
ORGANISMUnknown.
Unknown.

Unclassified.

1 (bases 1 to 3003)

Sloma,A. and Christianson,L.

Nucleic acids encoding a polypeptide having protease activity

Patent: US 5891701-A 41 06-APR-1999;

JOURNAL Location/Qualifiers

FEATURES

1..3003

/organism="unknown"

/mol_type="unassigned DNA"

ORIGIN

Alignment Scores:

Pred. No.: 8,41e-128 Length: 3003
Score: 2118.50 Matches: 405
Percent Similarity: 97.47% Conservative: 18
Best Local Similarity: 93.32% Mismatches: 10
Query Match: 94.37% Indels: 1
DB: 6 Gaps: 1

US-09-985-689A-2 (1-434) x AR069954 (1-3003)

Qy 1 AsnAspValAlaAlaArgGlyIleValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
Db 1470 AATGACGTGGCCGCTGCATTTGTGAAGCAGACGTCGCACAAATAAATTCCTTATAT 1529
Qy 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 1530 GGAACAGGACAGATTGTAGCAGTTGCTGATCTGGGCTTTGTATACAGAAAGAAATGACAGT 1589
Qy 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 1590 TCATGTCATGAAGCATTTCCGGGTAGATTACCGCATATATGCTGCTGGCAGAACGAT 1649
Qy 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
Db 1650 AACGCCAATGATCCAAATGGACATGGACATGCCATCTGTCTGGATCTGTGTAGAAAT--- 1706
Qy 81 AlaThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
Db 1707 GCTACAAATAAGGATGGCAGCCGCAAGCCATCTAGTCTTTCATCTATTATGGATAGT 1766
Qy 101 SerGlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaPheSer 120
Db 1767 GGTGGAGGGCTGGGAGGACTACCTGCTAATCTACAAACATTATTTCAGTCAAGCATATAGT 1826
Qy 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyrThrThr 140
Db 1827 GCTGGAGCGAGAAATTCATCGAATTCATGGGGGCTCCAGTAAACGCTGCCATATACGACA 1886
Qy 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
Db 1887 GACTCTCGAAATGTTGATGATTATGTGAGAAAAAATGATATGACGATCTTTTGGCGGC 1946
Qy 161 GlyAsnGluGlyProAsnGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
Db 1947 GGAATGTAGGAGCCCGGTAGCGGTACAATCAGTGCACCGAGAACACCAAAATATGGCATT 2006
Qy 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
Db 2007 ACAGTTGGGCAACCCGAAACCTACGTCCAAAGCTTCGGATCTTATCGGATATATTAAC 2066
Qy 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
Db 2067 CATGTTGCTCAATTCCTTACAGAGTCTCTACTAGATGGAGTATTAAGCCGAGCGCTC 2126
Qy 221 MetAlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
Db 2127 ATGGCACCAGGTACGTATATTTCTCTGCTAGATCATCATTAGTCTCCAGATTCCTCATTC 2186
Qy 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260

| | | | |
|------|----------|--|------|
| 2187 | TGGGCAAA | CCATGATAGTAAATATGCCTACATCGGTGGTACTTCTATGGCTACTCCAATT | 2244 |
| QY | 261 | ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValIysAsnArgGlyIleThrPro | 280 |
| Db | 2247 | GTAGCAGGTAATGTTGCACAAATTAAGGAGCAATTTGTGAAAAATAGAGGGTAACCTCT | 2306 |
| QY | 281 | LysProSerLeuLeuIysAlaIleLeuIleAlaGlyAlaAlaAspValGlyLeuGlyTyr | 300 |
| Db | 2307 | AAOCCTTCCTTTTAAAGAGCTGTTTAATTTGCAAGGTGCTCGGATGTGGACTTGGCTTT | 2366 |
| QY | 301 | ProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr | 320 |
| Db | 2367 | CCAAATGGTAACCAAGATGGGAAGAGTAACGTGTAGATAAATCCCTAAATGTCGCATTT | 2426 |
| QY | 321 | ValAsnGluSerSerAlaLeuSerThrSerGlnIysAlaThrTyrThrPheThrAlaThr | 340 |
| Db | 2427 | GTGAATGAACAGAGCCCTTTATCAACAAGTCAAAAGCAACAPATTCGTGTTACGGGTCAA | 2486 |
| QY | 341 | AlaGlyLysProLeuIysIleSerLeuValTrpSerAspAlaProIaSerThrThrAla | 360 |
| Db | 2487 | GCTGGTAACCCITTAAAAATATCACTTGTTTGGTCAGATGCACAGGTAGCAGCGGCA | 2546 |
| QY | 361 | SerValThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrArgTyr | 380 |
| Db | 2547 | TCACTAACTTTAGTGAATGATTTAGACTTTAGTAATCACTGCACAAATGAAGTAATAATC | 2606 |
| QY | 381 | ValGlyAsnAspPheSerAlaProPheAspAsnAsnTrpAspGlyArgAsnAsnValGlu | 400 |
| Db | 2607 | GTCCGAAATGACTTTTACAGCACCGTATGATAACAATTTGGGATGCGAGAAACAACGTGGAA | 2666 |
| QY | 401 | AsnValPheIleAsnSerProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn | 420 |
| Db | 2667 | AATGTGTTTATCAATGTCTCTCAAGCGGAACGTPATACGTGCAGGTGCAGGCTTACAAAT | 2726 |
| QY | 421 | ValProValGlyProGlnAsnPheSerLeuAlaIleValAsn | 434 |
| Db | 2727 | GTACACGTAAGTCCGCAAAACCTTTCTTTACCGATTGTACAT | 2768 |

RESULT 10
BD062155
LOCUS
BD062155 3003 bp DNA linear PAT 27-AUG-2002
Nucleic acids encoding a polypeptide having protease activity.
DEFINITION

| | |
|-----------|--|
| ACCESSION | BD062155 |
| VERSION | BD062155.1 |
| KEYWORDS | GI:22607760 |
| SOURCE | JP 200154529-A/39. |
| ORGANISM | Homo sapiens (human) |
| | Homo sapiens |
| | Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. |

REFERENCE 1 (bases 1 to 3003)
AUTHORS Sloma, A. and Christianson, L.
TITLE Nucleic acids encoding a polypeptide having protease activity
JOURNAL Patent: JP 2001514529-A 39 11-SEP-2001;

| NOVO NORDISK BIOTECH INC | COMMENT |
|---|---------|
| PN JP 2001514529-A/39 | |
| PD 11-SEP-2001 | |
| PF 09-JUN-1998 JP 1999503145 | |
| PR 12-JUN-1997 US 08/873479 | |
| PI ALAN SLOMA,LYNNE CHRISTIANSON | |
| PC C12N15/57,C12N15/75,C12N5/54,C12K14/00 | |
| CC Strandedness: Single; | |
| CC Topology: Linear; | |

```

FEATURES
  PH Key Location/Qualifiers.
  process: name: Location/Qualifiers
  1..3003
    /organism="Homo sapiens"
    /mol_type="genomic DNA"
    /db_xref="taxon:9606"
  source

```

ORIGIN

| | | | |
|-------------------|-----------|----------|------|
| Alignment Scores: | 8.41e-138 | Length: | 3003 |
| Pred. No.: | 2118.50 | Matches: | 405 |
| Score: | | | |

| Percent Similarity: | 97.47% | Conservative: | 18 |
|--|--------|--|------|
| Best Local Similarity: | 93.32% | Mismatches: | 10 |
| Query Match: | 94.37% | Indels: | 1 |
| DB: | 6 | Gaps: | 1 |
| US-09-585-689A-2 (1-434) x BD062155 (1-3003) | | | |
| Qy | 1 | AsnAspValAlaArgGlyLeuValysAlaAspValAlaGlnSerSerTyrGlyLeuTyr | 20 |
| Db | 1470 | AATGACGTGGCGTGGCGATTGTGAAGCAGACGTGCACAAATAACTTTGGCTTATAT | 1529 |
| Qy | 21 | GlyGlnGlyGlnIleValAlaAlaAspThrGlyLeuAspThrGlyArgAsnAspSer | 40 |
| Db | 1530 | GGACAGGRCAGATTGTGACAGTTGCTGATCTGGCTTGATACAGAAATGACAGT | 1589 |
| Qy | 41 | SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn | 60 |
| Db | 1590 | TCGATGCATGAAGCATTCGCGGTAGATTACCGCACTATATGCTCGGCAGAACGAAT | 1649 |
| Qy | 61 | AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly | 80 |
| Db | 1650 | AACGCCATGATCCAAATGGACATGGAAACCCATGTTGCTGGATCTGTGTAGAAAT--- | 1706 |
| Qy | 81 | AlaThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer | 100 |
| Db | 1707 | GCTACAAATAAAGGATGGCACGCAAGCAATCTAGTCTTTCAAATCTATTATGGATAGT | 1766 |
| Qy | 101 | SerGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaPheSer | 120 |
| Db | 1767 | GGTGGAGGGCTGGAGAGACTCTCTGCTTAATCTCAAAACATTATTCACTCAAGCATATAGT | 1826 |
| Qy | 121 | AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValaAsnGlyAlaTyrThrThr | 140 |
| Db | 1827 | GCTGGACGAGAAATTCATACGAATTCATGGGGGCTCCAGTAAACGGTGCTATACGACA | 1886 |
| Qy | 141 | AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla | 160 |
| Db | 1887 | GACTCTCGAAATGTGATGATTATGTAGAAAAAATGATAGCAGATCTTTTTCGGGCC | 1946 |
| Qy | 161 | GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle | 180 |
| Db | 1947 | GGAAATGAGGGACCCAGGTAGCGGTACAAATCAGTGCACCGAGAAACAGCAAAAATGCGATT | 2006 |
| Qy | 181 | ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn | 200 |
| Db | 2007 | ACAGTTGGGGCAACCGAAACCTACGTCCAAAGCTTCGGATCTTATCGGATAATATTAAAC | 2066 |
| Qy | 201 | HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal | 220 |
| Db | 2067 | CATGTTGCTCAATCTCTTCACGAGGTCTCTAGAGATGGAGCTATTAAAGCGGACGTC | 2126 |
| Qy | 221 | MetAlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe | 240 |
| Db | 2127 | ATGGCACCGGTACGTATATCTCTGCTAGATCATCATTAGCTCCAGATTTCTCATTC | 2186 |
| Qy | 241 | TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIle | 260 |
| Db | 2187 | TGGGCAACCATGATAGTAATAATGCCCTACATGGGTGCTCTTCTATGGCTACTCCAATT | 2246 |
| Qy | 261 | ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro | 280 |
| Db | 2247 | GTACGAGTAAATGTTGCACAAATTAAGGAGCATTTTGTGAAAAATAGAGGGGTAACTCCT | 2306 |
| Qy | 281 | LysProSerIleuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspValGlyLeuGlyTyr | 300 |
| Db | 2307 | AAGCCTTTCCCTTTTAAAGCTGCTTTAAATGTCAGGTGCTCGGATGTTGGACTTGGCTTT | 2366 |
| Qy | 301 | ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr | 320 |
| Db | 2367 | CCAAATGGTACCAAGGATGGGAAGAGTAACGTTAGATAAATCCCTAAATGTCGCATTT | 2426 |
| Qy | 321 | ValAsnGluSerSerAlaLeuSerThrSerGlnLysAlaThrTyrThrPheThrAlaThr | 340 |

```

Db      2427 GTGAATGAAACGAGCCCTTTATCAACAGTCACAAAGCAACATATTCGTTACCGGTCAA 2486
Qy      341 AlaGlyLysProLeuLysSerLeuValTTPSerAspAlaProAlaSerThrThraAla 360
Db      2487 GCTGTGAACCCCTTAAATAATACCTGTTGGTCAGATGCACGAGGTAGCAGCAGCGCA 2546
Qy      361 SerValThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrArgTyr 380
Db      2547 TCACAACTTTAGTGAATGATTAGACTTAGTAATCACTGCACCAATAGGAACATAATAC 2606
Qy      381 ValGlyAsnAspPheSerAlaProPheAspAsnAsnTTPAspGlyArgAsnAsnValGlu 400
Db      2607 GTCCGGAATGACTTTACAGCACCGGTATGATACCAATTCGGATGGCAGAAACAACTGGAA 2666
Qy      401 AsnValPheIleAsnSerProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
Db      2667 AATGTGTTTATCAATGCTCTCAAGCGGAAGGTATACAGTCGAGTGCAGGCTTACAT 2726
Qy      421 ValProValGlyProGlnAsnPheSerLeuAlaIleValAsn 434
Db      2727 GTACCAGTAAGTCGCAAAACCTTTCTTTAGCGATTGTACAT 2768

RESULT 11
AB046405 LOCUS          1299 bp      DNA      linear      BCT 23-JAN-2001
DEFINITION   Bacillus sp. SD521 PROD gene for protease, partial cds.
ACCESSION   AB046405
VERSION     AB046405.1  GI:12381942
KEYWORDS
SOURCE      Bacillus sp. SD521
ORGANISM    Bacillus sp. SD521
REFERENCE   1 (bases 1 to 1299)
AUTHORS     Saeki,K., Okuda,M., Hatada,Y., Kobayashi,T., Ito,S., Takami,H. and
            Horikoshi,K.
TITLE       Novel oxidatively stable subtilisin-like serine proteases from
            alkaliphilic Bacillus sp.: enzymatic properties, sequences, and
            evolutionary relationships
JOURNAL     Biochem. Biophys. Res. Commun. 279 (2), 313-319 (2000)
MEDLINE     20568675
PubMed     11118284
REFERENCE   2 (bases 1 to 1299)
AUTHORS     Saeki,K.
DIRECT SUBMISSION
SUBMITTED   (20-JUL-2000) Katsuhisa Saeki, Kao corporation,
            Biological Science Laboratory; Ichikaimachi Akabane 2606, Haga,
            Tochigi 321-3497, Japan (E-mail:387185@kantanet.kao.co.jp,
            Tel:81-285-68-7400, Fax:81-285-68-7403)
FEATURES
            Location/Qualifiers
            1..1299
                /organism="Bacillus sp. SD521"
                /mol_type="genomic DNA"
                /strain="SD521"
                /db_xref="taxon:133780"
            1..1299
                /gene="PROD"
                <1..>1299
                /gene="PROD"
                /codon_start=1
                /transl_table=11
                /product="protease"
                /protein_id="BAB21268.1"
                /db_xref="GI:12381943"
            /translation="NDVARGIVKADVQNNYGLYGGQGVAVADTGLDTRNDSSMHE
            AFRGKITALYALGRNTNNDPNHGTHVAGVLGNALNKGMAPQANLVFQSDSSGG
            LGLPSNLNTLFSQAWNAGARIHTNSWGPVNGVATNSROVDYVRNDMTVLFAAG
            NEGPNSTISAPATAKNAITVGCATENYPPSPGSLADPNHIAQSPSSGATPDGKPD
            VTAPGTFILSRSSLAPSSSPWANNYSKYAMGTGSMATPIVAGNVAQLREHFINKG
            ITPKPSLKAALAGATVDVGLGPGSDGWRVLDLSLVNVAIYNEATATAGTKATY
            SFQAGAKPLKISLWTDAPGSTASYTLSDTLVITPAENGKIVGNDFSPYDNNW
            DGRNVENVFINAQSGTYTIEVQAYNVPSGPQPSLAIVH"
ORIGIN

```

```

Alignment Scores:      2,88e-121      Length:      1299
Pred. No.:      2010.50      Matches:      384
Score:      94.47%      Conservatives:      26
Percent Similarity:      88.48%      Mismatches:      23
Best Local Similarity:      89.55%      Indels:      1
Query Match:      1      Gaps:      1
DB:      1
US-09-985-689A-2 (1-434) x AB046405 (1-1299)

Qy      1 AsnAspValAlaAlaArgGlyIleValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
Db      1 AATGATGTAGCAGAGGAATAGTAAAGCAGACGTTGGCAAAACAACTATACCGACTATAT 60
Qy      21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db      61 GGACAGGTCACAGTAGTTCAGTAGCGGATACGGGTTTAGATACAGGTCGTACCGTAGT 120
Qy      41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db      121 TCTATGCATGAGCATTCGCTGGGAAAATTACAGCTCTTTAGCGGTAGGAAGAACTAAC 180
Qy      61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
Db      181 AATGCAATGATCCGAATGGCAATGGTACGATGTAGCAGGTCTGTACTTGGTAAT--- 237
Qy      81 AlaThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
Db      238 GCTTTAATAAAGGAATGGCTCGCAAGCTAACTAGTCTTCCAATCTATTATGATAGC 297
Qy      101 SerGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaPheSer 120
Db      298 AGCGGAGGATTTAGTGATTTACCATCGAATTTGAATACGTTATTATTAGTCAGCTTGGAA 357
Qy      121 AlaGlyAlaArgIleHisThrAsnSerTyrGlyAlaAlaValAsnGlyAlaTyrThrThr 140
Db      358 GCTGGGGCTAGAAATTCATACTAACTCTTGGGGTGTCCAGTAATAGGAGCGTACACTGCT 417
Qy      141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
Db      418 AACTCGAGACAGTGGATGATGTGTGAAATAATGATATGATGATGATGATGATGATGATG 477
Qy      161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
Db      478 GGTAATAGAGGTCTTAATTCAGGAACAAATTAGTGCTCCAGGCACACGCAAAATGCAAT 537
Qy      181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
Db      538 ACGGTCGGCGCAACGGAAAACCTATCCCGGAGCTTCGGTTCATTAGCAGATAACCCAAAT 597
Qy      201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
Db      598 CATATCGCACAAATTTTCATCAAGAGGAGCTACGAGAGATGGAGCAATTAACACGACGTA 657
Qy      221 MetAlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
Db      658 ACAGCTCCTGGAACATTTATTTTATCAGCACGTCCTTCCTTAGCCCCAGACACTTCGTTT 717
Qy      241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260
Db      718 TGGCGGAAATTAACAGTAGTAGTATGGCTACATGGCGGCTACTCTTAGCGCACCTATA 777
Qy      261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
Db      778 GTTTCGGGGAATTCGCGCAATTTACCGAGCATTTTATAAAAAATAGAGGAATTAACCT 837
Qy      281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspValGlyLeuGlyTyr 300
Db      838 AAACCTTCCTTAATAAAGCTGACCTTATCGCTGGGGCTACTCGTGCTGGGCTAGGATAT 897
Qy      301 ProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
Db      898 CCAAGTGTGTGACCAACGCTGGGGCGCTGTTACTCTAGATAAATCGTTAAATGTAGCGTAT 957

```


QY 301 ProAsnGlyVasGlnGlyTrrpGlyVargValThrLeuAspLysSerLeuAsnValAlaThr 320
 Db 898 CCAAGTGGTACCAAGCTGGGGGGTGTACTTTAGATATAATCTTAAATGTAGCGTAT 957
 QY 321 ValAsnGluSerSerAlaLeuSerThrSerGlnLysAlaThrTyrThrPheThrAlaThr 340
 Db 958 GTCAATGAAGCAACTGCATTAACAACAGCAAAAAAGCAAGCTATTCGTTCCAAAGCAA 1017
 QY 341 AlaGlyLysProLeuValSerLeuValTrrpSerAspAlaProAlaSerThrThrAla 360
 Db 1018 GCGGGTAAACCAATTAAATCTCGTTAGTAGGACAGATGCACCTGGGAAGTACACAGCA 1077
 QY 361 SerValThrLeuValAsnAspLeuValThrAlaProAsnGlyThrArgTyr 380
 Db 1078 TCTTATACACTAGTTAATGATTAGATCTAGTTACTTCTGCTCCGAATGGACAAAATAT 1137
 QY 381 ValGlyVasAspPheSerAlaProPheAspAsnThrAspGlyVargAsnValGlu 400
 Db 1138 GTAGTAATGATTTTGTAGTTATCCCTATGATATAATTCGGATGGTTCGCAAAATGTTGAG 1197
 QY 401 AsnValPheLeuAsnSerProGlnSerGlyThrTyrThrThrLeuValGlnAlaThrAsn 420
 Db 1198 AACGTATTATAAACGCTCCGCAATCTGGAACGTATACAAATGAGTTCAAGCGTATPAC 1257
 QY 421 ValProValGlyProGlnAsnPheSerLeuAlaLysValAsn 434
 Db 1258 GTTCCATCTGGACACAGCGTTCTCACTAGTATCGTACAT 1299

RESULT 13
 AB046404
 LOCUS AB046404 1299 bp DNA linear BCT 23-JAN-2001
 DEFINITION Bacillus sp. Y PROC gene for protease, partial cds.
 ACCESSION AB046404
 VERSION AB046404.1 GI:12381940
 KEYWORDS
 SOURCE
 ORGANISM
 Bacillus sp. Y
 Bacillus sp. Y
 Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 REFERENCE
 AUTHORS
 Horikoshi,K.
 TITLE Novel oxidatively stable subtilisin-like serine proteases from alkaliphilic Bacillus spp.: enzymatic properties, sequences, and evolutionary relationships
 JOURNAL Biochem. Biophys. Res. Commun. 279 (2), 313-319 (2000)
 MEDLINE 20568675
 PUBMED 11118284
 REFERENCE
 AUTHORS
 Saeki,K.
 TITLE Direct Submission
 JOURNAL Submitted (20-JUL-2000) Katsuhisa Saeki, Kao corporation, Biological Science Laboratory, Ichikaimachi Akabane 2606, Haga, Tochigi 321-3497, Japan (E-mail:387185@kaxanet.kao.co.jp, Tel:81-285-68-7400, Fax:81-285-68-7403)
 FEATURES
 Location/Qualifiers
 1..1299
 /organism="Bacillus sp. Y"
 /mol_type="genomic DNA"
 /strain="Y"
 /db_xref="taxon:133779"
 1..1299
 /gene="PROC"
 <1..>1299
 /gene="PROC"
 /codon_start=1
 /evidence=experimental
 /transl_table=11
 /product="protease"
 /protein_id="BAB21267.1"
 /db_xref="GI:12381941"
 /translation="NDVARGIVKADVQNNGLYGGQGVAVADTGLDITGRNDSMEH
 AFRGKITALYALGRNNASDPNGHGTAVAGSVLGNALNKGWAPQANLVFQSIIMDSG3
 LGGLPSNLNTLPSQAWNAGARLHTNSWCAPVNGAYTANRSQVDEYVNNNDTVLFAAG

NEGSPSTISAPCTAKNAITGATENYRPSFGSIADNPNHIAQFSSRGATRDGRKPD
 VTAPGTFILUSASLAPDSFWANTNSKTAINGGTSMATPIVAGNVATQALREHFKNG
 ITPKSLIKALIIAGATDVLGYPNGDQWVTLNLSNVAYNEATALATGQKATY
 SFOAQAGKPLKISLVMTDAPSTTAYTLVNDLDELVTAPNGKQYVNDPSFYDNNW
 DGRNNVNFVINAPOSQGTIIIEVQAYNVPSGPFQFSLAIVH"
 ORIGIN
 Alignment Scores:
 Pred. No.: 1 48e-120 Length: 1299
 Score: 1999.50 Matches: 382
 Percent Similarity: 94.24% Conservative: 27
 Best Local Similarity: 88.02% Mismatches: 24
 Query Match: 89.06% Indels: 1
 DB: 1 Gaps: 1
 US-09-985-689a-2 (1-434) x AB046404 (1-1299)
 QY 1 AsnAspValAlaArgGlyLeuValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
 Db 1 AATGATGTAGCAGAGGGATAGTAAAGCTGTGTGCACAAAACAATACGGATTATAT 60
 QY 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyVargAsnAspSer 40
 Db 61 GGCACAGGTCAGTAGTTGTCAGTAGCGGACACAGCTTAGATACAGGTCGTACGATAT 120
 QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
 Db 121 TCTATGCATGACGATTCGCGGGAATACACAGCTCTTTACGCGTTAGGAGAACTAAT 180
 QY 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
 Db 181 AATCGGAGTATCCGAAATCGGCATCGCACACATGTAGCAGGTTCTGTACTTGGTAAT --- 237
 QY 81 AlaThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
 Db 238 GCITTTAAATAAGGAATGCTCCGCAAGCTAACTAGTCTTCCAACTATTATGGATAGC 297
 QY 101 SerGlyGlyLeuGlyClyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaPheSer 120
 Db 298 AGCGGAGGATAGTGGCTTACCATCGAACTAAATACGTTATTAGTCAAGCTTGAAT 357
 QY 121 AlaGlyAlaArgIleHisThrAsnSerTrrpGlyAlaAlaValAsnGlyAlaTyrThrThr 140
 Db 358 GCTGGAGCAAGAAATTCATCACTAACTCTTGGGAGCCCAAGTAATCGACGTACACTGCT 417
 QY 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaLys 160
 Db 418 AACTCGAGACAAGTGGATGAATATGTTCGAAATAATGATATACCGTACTTTTTCGAGCT 477
 QY 161 GlyAsnGluGlyProAsnGlyThrIleSerAlaProGlyThrAlaLysAsnAlaLys 180
 Db 478 GGTAAATGAAGTCTTAATTCAGGAACAATAGTCTCCAGGTACAGCGAAAATGCTATT 537
 QY 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
 Db 538 ACGGTCGGCGCAACGGAAAACTATCCCAAGCTTCGGTTCGATAGCAGATACCCAAAT 597
 QY 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
 Db 598 CATATTGCACAATTTTCATCGAGAGAGCTACGAGGATGGACGATTAAGCTGACGTA 657
 QY 221 MetAlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
 Db 658 ACAGCTCTCTGGAACATTTATTTATCAGCACGTTCTTCTCTAGTCCAGACTCTTCGTTT 717
 QY 241 TrpAlaAsnHisAspSerLysThrAlaTyrMetGlyThrSerMetAlaThrProIle 260
 Db 718 TGGCGGAATTTATAACAGTAAATACGGTATATGGCGGGTACCTCCATGCGGACACCTATT 777
 QY 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
 Db 778 GTTGCAGGGAATGTCGCGCAATTACGTGAGCATTTTATATAAAAAATAGAGGTATTACTCTCT

```
QY 281 LysProSerLeuLeuLysAlaLeuLeuAlaGlyValAlaAspValGlyLeuGlyTyr 300
Db 838 AAGCTTCTTTATAAAGCTGACCTTATCCCTGGTCTACTGATGTTTGTAGATAT 897
QY 301 ProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
Db 898 CCTAATGGTGCACCAAGCTGGGGCGGTGTACTCTAAATAAATCGTTAAATCGTAT 957
QY 321 ValAsnGlySerAlaLeuSerThrSerGlnLysAlaThrTyrThrPheThrAlaThr 340
Db 958 GTCAATGACCACTGATGATAGCCACGACGACAAAGCAACGATTCGTTCACGACAA 1017
QY 341 AlaGlyLysProLeuLysLeuValThrSerAspAlaProAlaSerThrThrAla 360
Db 1018 GCGGGTAAACCTTTAAATCTCGTTAGTAGCAGACAGATGCTCTCGGAATGCAACTGCA 1077
QY 361 SerValThrLeuValAsnAspLeuValThrAlaProAsnGlyThrArgTyr 380
Db 1078 TCTTATACACTAGTTTATGATTTAGATCTAGTTTACTGCTCCGAAATGGACAAATAT 1137
QY 381 ValGlyAsnAspPheSerAlaProPheAspAsnAsnTyrAspGlyArgAsnValGlu 400
Db 1138 GTAGGAATGATTTAGTTATCTTATGATATAAATGCTGGATGCTCGCAACAATGTTGAG 1197
QY 401 AsnValPheLeuAsnSerProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
Db 1198 AAGCTATTTATAAACGCTCGCAATCTGGAACGATATTAATTGAGGTTCAAGCGTATAAT 1257
QY 421 ValProValGlyProGlnAsnPheSerLeuAlaIleValAsn 434
Db 1258 GTACCATCTGCCCCACAGCGTTTCTCACTAGCTATCGTACAT 1299

RESULT 14
E03808
LOCUS      E03808      2218 bp      DNA      linear      PAT 29-SEP-1997
DEFINITION DNA encoding alkaline protease Ya.
ACCESSION E03808
VERSION    E03808.1 GI:2172022
KEYWORDS   JP 1992197182-A/1.
SOURCE     Bacillus sp.
ORGANISM   Bacillus sp.
REFERENCE  1 (bases 1 to 2218)
AUTHORS    Tobe,S., Odera,M. and Asai,Y.
TITLE      DNA CODING ALKALINE PROTEASE YA ENZYME AND PRODUCTION OF ALKALINE
           PROTEASE YA USING THE DNA
JOURNAL    Patent: JP 1992197182-A 1 16-JUL-1992;
           LEON CORP.
COMMENT    OS Bacillus sp.
           PN JP 1992197182-A/1
           PD 16-JUL-1992
           PF 28-NOV-1990 JP 1990327110
           PI TOBE SEIICHI, ODERA MOTOKYASU, ASAI YOSHIO
           PC C12N15/57,C12N15/386,C12N9/54, (C12N15/57,C12R1/07), (C12N9/54,
           PC C12R1/07);
           CC strandedness: Double;
           CC topology: Linear;
           CC hypothetical: No;
           CC anti-sense: No;
           CC *source: strain=Y strain;
           CC *source: clone=pUB8A;
           FH Key Location/Qualifiers
           FT CDS 218..2125
           FT /product='precursor of alkaline protease Ya'
           FT sig_peptide 218..823
           FT mat_peptide 824..2122
           FT /product='Precursor of alkaline protease Ya'.
           FT Location/Qualifiers
           FT 1..2218
           FT /organism='Bacillus sp.'
           FT /mol_type='genomic DNA'
           FT /db_xref='taxon:1409'
```

```
ORIGIN
Alignment Scores: 3.33e-120 Length: 2218
Pred. No.: 1998.50 Matches: 382
Score: 94.24% Conservative: 27
Best Local Similarity: 88.02% Mismatches: 24
Query Match: 89.02% Indels: 1
DB: 6 Gaps: 1

US-09-985-689A-2 (1-434) x E03808 (1-2218)
QY 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
Db 824 AATGATGTAGCAGAGGAGTAGTAAAGCTGATGTTGCAAAACAACTACCGATTATAT 883
QY 21 GlyGlnGlyGlnIleValAlaAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 884 GGACAAGGTCACTGATGTTGCGTAGCGACACAGGCTTAGATACAGTCTGTAACGATAGT 943
QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 944 TCTATGATGAAGCATTCGCGGGAATAATCACAGCTCTTTACGCTTAGGAGAACTAAT 1003
QY 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
Db 1004 AATCGAGTGTATCGAATGGGCATGGCACATGTAGCAGGTCTGTACTTGGTAAT--- 1060
QY 81 AlaThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
Db 1061 GCTTTAAATAAAGGAATGGCTCCGCAAGCTAACTAGTCTTCCAATCTATTATGATAGC 1120
QY 101 SerGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaPheSer 120
Db 1121 AGCGGAGGATTAGTGCTGCTTACCATCGAATTAATAACGTATTATTAGTCAAGCTTGAAT 1180
QY 121 AlaGlyAlaArgIleHisThrAsnSerTyrGlyValAlaAlaValAsnGlyAlaTyrThr 140
Db 1181 GCTGAGCAGAAATTCATACTACTCTTGGGAGCCCACTAATGAGGAGTACACTGCT 1240
QY 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
Db 1241 AACTCGAGACAAGTGGATGATGTTGCAAAATAATGATATGACGCTACTTTTGGAGCT 1300
QY 161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
Db 1301 GGTATGAGGTCTCTAATTCAGGACAAATAGTGTCTCCAGGTACAGCGAATAATGCTATT 1360
QY 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
Db 1361 ACGGTCCGCGCACACGAAACTATCGCCCAAGCTTCGGTTTCGATAGCAGATAACCCAAAT 1420
QY 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
Db 1421 CATATTGCACAAATTTTATTCGAGGAGGAGCTACGAGGAGTGGACGAATTAAGCTGACGTA 1480
QY 221 MetAlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
Db 1481 ACAGCTCTCGAACAATTTATTTATCAGCACGTTCTTCTTAGCTCCAGACTCTCTCGTTT 1540
QY 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIle 260
Db 1541 TGGCGGAATTAACAGTAATAACGCTATATGCGCGGTACCTCCATGGGACACCTATT 1600
QY 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
Db 1601 GTTCAGGGAATGTCGGCAATTAACGTGAGCAATTTATAAAAAATAGAGGTATTACTCCT 1660
QY 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyValAlaAspValGlyLeuGlyTyr 300
Db 1661 AAGCCTCTTTTAAATAAAGCTGCACTTATCGTGGTCTACTGATGTGTGTTAGGATAT 1720
QY 301 ProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
```

```
Db 1721 CCTAGTGTGACCAAGCTGGGGCGTGTACTCTAGATAAATCGTAAATGTAGCGTAT 1780
Qy 321 ValAenGluSerSerAlaLeuSerThrSerGlnLysAlaThrTyrThrPheThrAlaThr 340
Db 1781 GTCATGAAGCAACTGCATTAGCCACAGCAGCAAAAGCAACGTATTCGTTCCAGACAA 1840
Qy 341 AlaGlyLysProLeuLysLysSerLeuValThrPheSerAspAlaProAlaSerThrAla 360
Db 1841 CGGGTAACCTTTAAATCTCGTTAGTAGGACAGATGCTCTCGAAGTACAACTGCA 1900
Qy 361 SerValThrLeuValAsnAspLeuValThrAlaProAsnGlyThrArgTyr 380
Db 1901 TCTATACACAGTAAATGATTAGATCTAGTATTACTGCTCGAATGGACAAATAT 1960
Qy 381 ValGlyAsnAspPheSerAlaProPhePheAsnAsnThrAspGlyArgAsnValGlu 400
Db 1961 GTAGAAATGATTTTGTATCTTATGATATTAACCTGGATGCTCGCAACATGTTGAG 2020
Qy 401 AsnValPheLeuAsnSerProGlnSerGlyThrTyrThrLeuGluValGlnAlaTyrAsn 420
Db 2021 AACGTATTTTAAACGCTCCGCAATCTGGACGTATATTAATGAGTTCAAGCGTATAT 2080
Qy 421 ValProValGlyProGlnAsnPheSerLeuAlaLeuValAsn 434
Db 2081 GTACCATCTGGCCACAGCGTTCTCTCACTAGTATCGTACAT 2122

RESULT 15
AF268611/c
LOCUS
DEFINITION
  uncultured marine group II euryarchaeote BBAC37F11, BAC clone
ACCESSION
  AF268611
VERSION
  AF268611.1 GI:9664575
KEYWORDS
  .
SOURCE
  uncultured marine group II euryarchaeote 37F11
  Archaea; Euryarchaeota; Marine Group II; environmental samples.
REFERENCE
  1 (bases 1 to 60006)
  Nguyen,L.P., Villacorta,R., Amjadi,M., Garrigues,C.,
  Jovanovich,S.B., Feldman,R.A. and Delong,E.F.
  Construction and analysis of bacterial artificial chromosome
  libraries from a marine microbial assemblage
  Environ. Microbiol. 2 (5), 516-529 (2000)
  2 (bases 1 to 60006)
  Beja,O., Suzuki,M.T., Koonin,E.V., Aravind,L., Hadd,A.,
  Nguyen,L.P., Villacorta,R., Amjadi,M., Garrigues,C.,
  Jovanovich,S.B., Feldman,R.A. and Delong,E.F.
  Direct Submission
  Submitted (17-MAY-2000) R & D, Monterey Bay Aquarium Research
  Institute, P.O. Box 628, Moss Landing, CA 95039-0628, USA
  Location/Qualifiers
    1..60006
    /organism="uncultured marine group II euryarchaeote 37F11"
    /mol_type="genomic DNA"
    /db_xref="taxon:133822"
    /clone="BAC 37F11"
    /note="environmental sample"
    1484..3178
    /note="37F11#1"
    /codon_start=1
    /evidence=not experimental
    /transl_table=11
    /product="S-layer protein"
    /protein_id="AAF97179.1"
    /db_xref="GI:9664576"
    /translation="MEPGQTELDIFVQSPSSATAGQNVIVKIRVSGSGGLEPEPI
    PVVSPSFFNLASQGVSEVGGYPLANNQNTWNTSHVTLENLPGWNISGMQ
    SASPSFHARGIPFEIIPPTWQKPLELSIVHPQGMVESITVIRHNSFASFP
    VLSSGAGSQSPSLHNSINPSTVTVPDATNQDEYRLIEBGINFETISTNDIAIQ

CDS
  1..60006
  /product="S-layer protein"
  /protein_id="AAF97179.1"
  /db_xref="GI:9664576"
  /translation="MEPGQTELDIFVQSPSSATAGQNVIVKIRVSGSGGLEPEPI
  PVVSPSFFNLASQGVSEVGGYPLANNQNTWNTSHVTLENLPGWNISGMQ
  SASPSFHARGIPFEIIPPTWQKPLELSIVHPQGMVESITVIRHNSFASFP
  VLSSGAGSQSPSLHNSINPSTVTVPDATNQDEYRLIEBGINFETISTNDIAIQ
```

```

CDS
  3223..3891
  /note="37F11#2"
  /codon_start=1
  /evidence=not experimental
  /transl_table=11
  /product="unknown"
  /protein_id="AAF97180.1"
  /db_xref="GI:9664577"
  /translation="MNQSNSTSAFYTLDDGCVENLDETFOALLIIRGNHQLWEKLV
  NREASTRIUKQDEIDDRNKITWDDATLMTFCVRLNRSAPIMLGNKRRFGR
  PHHTQSQYILENLIIPSQSEGLYSLDEHLEYLTERSEMNTHKEFLKAGL
  MLGTYITFEVKALRSMILGSGMWVSREPLDGGVREAVRHLNALLMAERRGAGLIH
  ROHE"
  complement(3892..4797)
  /note="37F11#3"
  /codon_start=1
  /evidence=not experimental
  /transl_table=11
  /product="SAICAR synthase"
  /protein_id="AAF97181.1"
  /db_xref="GI:9664578"
  /translation="MGEMLYQGVKQVMSTDDPDLIERFTNQISVFDQIIPSLIPRK
  GSLNRTTAHFKLYEQSGVCGTHLVELNAPRCLVRKVVEVKEPGALPRDAEWFVP
  LFIIVHYLUSGAWRRFORGETLBSLVGSDAQEYGVLSSEFVEVTTTFRADFNRID
  REAQISNITQDFDSIVQAALKYDEIEREAKNGLIHVDGKKEFALGPGKRVLY
  DTFTLDEDRWDAAEYNGECIELSEKFEVTRTHYNSGHAELKLARDSGAIDPPTPA
  LPSEVIEETASLYSSMYERLTSGQP"
  complement(4846..5508)
  /note="37F11#4"
  /codon_start=1
  /evidence=not experimental
  /transl_table=11
  /product="translation initiation factor 6"
  /protein_id="AAF97182.1"
  /db_xref="GI:9664579"
  /translation="MPIDTLDHGSDQIGIHLSVGNVLFHPRELPTTIEKLEOVVG
  LELAPISGGSNLIGALLAGNKGNAVADIATESIDALTAYGVVMEGGVNTAGNL
  LVNQGCVGSPSIDGLDILSELVGLDITSTIGGQDVGSLGVVNSQGVLLHPDV
  APEVLLIEILGVPPMVGTGVSFPGYVAGICASNNGALASGETTGPENLRIEDALG
  LI"
  5597..5745
  /product="5S ribosomal RNA"
  5816..8611
  /product="23S ribosomal RNA"
  complement(8888..10192)
  /note="37F11#7; contains Zn-ribbon and TPR-like repeats"
  /codon_start=1
  /evidence=not experimental
  /transl_table=11
  /product="unknown"
  /protein_id="AAF97183.1"
  /db_xref="GI:9664580"
  /translation="MNECAVCGILYLAGSPCAGCQSLRAQOQSFSDSDAALPTEVPG
  LIDAAQWYDLEGIKAPVEEAAASLPFGGSESTNISRLEPFGISHPEGIIPFEL
  SNQDSQEEEVQHVPEVIESTNLEPVVALETEQDSPPVVSASVEIEPVLAPVAPD
  LPLIEBPVQVPPVPGPIRLTAIPFVEPTVDTDLNENWISIESFGGSDQIIFTAEDVIE
  VVFSLEDIVHVEHDGENYPOWEKDSVYSEKSLQFPLHPAQALMVDVGTDSARAL
  LESGPTSIGGQWRSAASFOQIVARPDQSGAMNNGYGTALLQVAVMWSNDPIEVS
  NSATQPEAAILSRVNRINSQSPESINLSQALLISGEEKALLDMAGNESQRES
  NFNTRAAILAQGYDEARLVLVQADKLASGNLLKLAL"
  complement(10248..11969)
  /note="37F11#8"
  /codon_start=1
  /evidence=not experimental
  /transl_table=11
  /product="permease"
  /protein_id="AAF97184.1"
  /db_xref="GI:9664581"
  /translation="MNECAVCGILYLAGSPCAGCQSLRAQOQSFSDSDAALPTEVPG
  LIDAAQWYDLEGIKAPVEEAAASLPFGGSESTNISRLEPFGISHPEGIIPFEL
  SNQDSQEEEVQHVPEVIESTNLEPVVALETEQDSPPVVSASVEIEPVLAPVAPD
  LPLIEBPVQVPPVPGPIRLTAIPFVEPTVDTDLNENWISIESFGGSDQIIFTAEDVIE
  VVFSLEDIVHVEHDGENYPOWEKDSVYSEKSLQFPLHPAQALMVDVGTDSARAL
  LESGPTSIGGQWRSAASFOQIVARPDQSGAMNNGYGTALLQVAVMWSNDPIEVS
  NSATQPEAAILSRVNRINSQSPESINLSQALLISGEEKALLDMAGNESQRES
  NFNTRAAILAQGYDEARLVLVQADKLASGNLLKLAL"
  complement(10248..11969)
  /note="37F11#8"
  /codon_start=1
  /evidence=not experimental
  /transl_table=11
  /product="permease"
  /protein_id="AAF97184.1"
  /db_xref="GI:9664581"
  /translation="MNECAVCGILYLAGSPCAGCQSLRAQOQSFSDSDAALPTEVPG
  LIDAAQWYDLEGIKAPVEEAAASLPFGGSESTNISRLEPFGISHPEGIIPFEL
  SNQDSQEEEVQHVPEVIESTNLEPVVALETEQDSPPVVSASVEIEPVLAPVAPD
  LPLIEBPVQVPPVPGPIRLTAIPFVEPTVDTDLNENWISIESFGGSDQIIFTAEDVIE
  VVFSLEDIVHVEHDGENYPOWEKDSVYSEKSLQFPLHPAQALMVDVGTDSARAL
  LESGPTSIGGQWRSAASFOQIVARPDQSGAMNNGYGTALLQVAVMWSNDPIEVS
  NSATQPEAAILSRVNRINSQSPESINLSQALLISGEEKALLDMAGNESQRES
  NFNTRAAILAQGYDEARLVLVQADKLASGNLLKLAL"
  complement(10248..11969)
  /note="37F11#8"
  /codon_start=1
  /evidence=not experimental
  /transl_table=11
  /product="permease"
  /protein_id="AAF97184.1"
  /db_xref="GI:9664581"
  /translation="MNECAVCGILYLAGSPCAGCQSLRAQOQSFSDSDAALPTEVPG
  LIDAAQWYDLEGIKAPVEEAAASLPFGGSESTNISRLEPFGISHPEGIIPFEL
  SNQDSQEEEVQHVPEVIESTNLEPVVALETEQDSPPVVSASVEIEPVLAPVAPD
  LPLIEBPVQVPPVPGPIRLTAIPFVEPTVDTDLNENWISIESFGGSDQIIFTAEDVIE
  VVFSLEDIVHVEHDGENYPOWEKDSVYSEKSLQFPLHPAQALMVDVGTDSARAL
  LESGPTSIGGQWRSAASFOQIVARPDQSGAMNNGYGTALLQVAVMWSNDPIEVS
  NSATQPEAAILSRVNRINSQSPESINLSQALLISGEEKALLDMAGNESQRES
  NFNTRAAILAQGYDEARLVLVQADKLASGNLLKLAL"
  complement(10248..11969)
  /note="37F11#8"
  /codon_start=1
  /evidence=not experimental
  /transl_table=11
  /product="permease"
  /protein_id="AAF97184.1"
  /db_xref="GI:9664581"
  /translation="MNECAVCGILYLAGSPCAGCQSLRAQOQSFSDSDAALPTEVPG
  LIDAAQWYDLEGIKAPVEEAAASLPFGGSESTNISRLEPFGISHPEGIIPFEL
  SNQDSQEEEVQHVPEVIESTNLEPVVALETEQDSPPVVSASVEIEPVLAPVAPD
  LPLIEBPVQVPPVPGPIRLTAIPFVEPTVDTDLNENWISIESFGGSDQIIFTAEDVIE
  VVFSLEDIVHVEHDGENYPOWEKDSVYSEKSLQFPLHPAQALMVDVGTDSARAL
  LESGPTSIGGQWRSAASFOQIVARPDQSGAMNNGYGTALLQVAVMWSNDPIEVS
  NSATQPEAAILSRVNRINSQSPESINLSQALLISGEEKALLDMAGNESQRES
  NFNTRAAILAQGYDEARLVLVQADKLASGNLLKLAL"
  complement(10248..11969)
  /note="37F11#8"
  /codon_start=1
  /evidence=not experimental
  /transl_table=11
  /product="permease"
  /protein_id="AAF97184.1"
  /db_xref="GI:9664581"
  /translation="MNECAVCGILYLAGSPCAGCQSLRAQOQSFSDSDAALPTEVPG
  LIDAAQWYDLEGIKAPVEEAAASLPFGGSESTNISRLEPFGISHPEGIIPFEL
  SNQDSQEEEVQHVPEVIESTNLEPVVALETEQDSPPVVSASVEIEPVLAPVAPD
  LPLIEBPVQVPPVPGPIRLTAIPFVEPTVDTDLNENWISIESFGGSDQIIFTAEDVIE
  VVFSLEDIVHVEHDGENYPOWEKDSVYSEKSLQFPLHPAQALMVDVGTDSARAL
  LESGPTSIGGQWRSAASFOQIVARPDQSGAMNNGYGTALLQVAVMWSNDPIEVS
  NSATQPEAAILSRVNRINSQSPESINLSQALLISGEEKALLDMAGNESQRES
  NFNTRAAILAQGYDEARLVLVQADKLASGNLLKLAL"
  complement(10248..11969)
  /note="37F11#8"
  /codon_start=1
  /evidence=not experimental
  /transl_table=11
  /product="permease"
  /protein_id="AAF97184.1"
  /db_xref="GI:9664581"
  /translation="MNECAVCGILYLAGSPCAGCQSLRAQOQSFSDSDAALPTEVPG
  LIDAAQWYDLEGIKAPVEEAAASLPFGGSESTNISRLEPFGISHPEGIIPFEL
  SNQDSQEEEVQHVPEVIESTNLEPVVALETEQDSPPVVSASVEIEPVLAPVAPD
  LPLIEBPVQVPPVPGPIRLTAIPFVEPTVDTDLNENWISIESFGGSDQIIFTAEDVIE
  VVFSLEDIVHVEHDGENYPOWEKDSVYSEKSLQFPLHPAQALMVDVGTDSARAL
  LESGPTSIGGQWRSAASFOQIVARPDQSGAMNNGYGTALLQVAVMWSNDPIEVS
  NSATQPEAAILSRVNRINSQSPESINLSQALLISGEEKALLDMAGNESQRES
  NFNTRAAILAQGYDEARLVLVQADKLASGNLLKLAL"
  complement(10248..11969)
  /note="37F11#8"
  /codon_start=1
  /evidence=not experimental
  /transl_table=11
  /product="permease"
  /protein_id="AAF97184.1"
  /db_xref="GI:9664581"
  /translation="MNECAVCGILYLAGSPCAGCQSLRAQOQSFSDSDAALPTEVPG
  LIDAAQWYDLEGIKAPVEEAAASLPFGGSESTNISRLEPFGISHPEGIIPFEL
  SNQDSQEEEVQHVPEVIESTNLEPVVALETEQDSPPVVSASVEIEPVLAPVAPD
  LPLIEBPVQVPPVPGPIRLTAIPFVEPTVDTDLNENWISIESFGGSDQIIFTAEDVIE
  VVFSLEDIVHVEHDGENYPOWEKDSVYSEKSLQFPLHPAQALMVDVGTDSARAL
  LESGPTSIGGQWRSAASFOQIVARPDQSGAMNNGYGTALLQVAVMWSNDPIEVS
  NSATQPEAAILSRVNRINSQSPESINLSQALLISGEEKALLDMAGNESQRES
  NFNTRAAILAQGYDEARLVLVQADKLASGNLLKLAL"
  complement(10248..11969)
  /note="37F11#8"
  /codon_start=1
  /evidence=not experimental
  /transl_table=11
  /product="permease"
  /protein_id="AAF97184.1"
  /db_xref="GI:9664581"
  /translation="MNECAVCGILYLAGSPCAGCQSLRAQOQSFSDSDAALPTEVPG
  LIDAAQWYDLEGIKAPVEEAAASLPFGGSESTNISRLEPFGISHPEGIIPFEL
  SNQDSQEEEVQHVPEVIESTNLEPVVALETEQDSPPVVSASVEIEPVLAPVAPD
  LPLIEBPVQVPPVPGPIRLTAIPFVEPTVDTDLNENWISIESFGGSDQIIFTAEDVIE
  VVFSLEDIVHVEHDGENYPOWEKDSVYSEKSLQFPLHPAQALMVDVGTDSARAL
  LESGPTSIGGQWRSAASFOQIVARPDQSGAMNNGYGTALLQVAVMWSNDPIEVS
  NSATQPEAAILSRVNRINSQSPESINLSQALLISGEEKALLDMAGNESQRES
  NFNTRAAILAQGYDEARLVLVQADKLASGNLLKLAL"
  complement(10248..11969)
  /note="37F11#8"
  /codon_start=1
  /evidence=not experimental
  /transl_table=11
  /product="permease"
  /protein_id="AAF97184.1"
  /db_xref="GI:9664581"
  /translation="MNECAVCGILYLAGSPCAGCQSLRAQOQSFSDSDAALPTEVPG
  LIDAAQWYDLEGIKAPVEEAAASLPFGGSESTNISRLEPFGISHPEGIIPFEL
  SNQDSQEEEVQHVPEVIESTNLEPVVALETEQDSPPVVSASVEIEPVLAPVAPD
  LPLIEBPVQVPPVPGPIRLTAIPFVEPTVDTDLNENWISIESFGGSDQIIFTAEDVIE
  VVFSLEDIVHVEHDGENYPOWEKDSVYSEKSLQFPLHPAQALMVDVGTDSARAL
  LESGPTSIGGQWRSAASFOQIVARPDQSGAMNNGYGTALLQVAVMWSNDPIEVS
  NSATQPEAAILSRVNRINSQSPESINLSQALLISGEEKALLDMAGNESQRES
  NFNTRAAILAQGYDEARLVLVQADKLASGNLLKLAL"
  complement(10248..11969)
  /note="37F11#8"
  /codon_start=1
  /evidence=not experimental
  /transl_table=11
  /product="permease"
  /protein_id="AAF97184.1"
  /db_xref="GI:9664581"
  /translation="MNECAVCGILYLAGSPCAGCQSLRAQOQSFSDSDAALPTEVPG
  LIDAAQWYDLEGIKAPVEEAAASLPFGGSESTNISRLEPFGISHPEGIIPFEL
  SNQDSQEEEVQHVPEVIESTNLEPVVALETEQDSPPVVSASVEIEPVLAPVAPD
  LPLIEBPVQVPPVPGPIRLTAIPFVEPTVDTDLNENWISIESFGGSDQIIFTAEDVIE
  VVFSLEDIVHVEHDGENYPOWEKDSVYSEKSLQFPLHPAQALMVDVGTDSARAL
  LESGPTSIGGQWRSAASFOQIVARPDQSGAMNNGYGTALLQVAVMWSNDPIEVS
  NSATQPEAAILSRVNRINSQSPESINLSQALLISGEEKALLDMAGNESQRES
  NFNTRAAILAQGYDEARLVLVQADKLASGNLLKLAL"
  complement(10248..11969)
  /note="37F11#8"
  /codon_start=1
  /evidence=not experimental
  /transl_table=11
  /product="permease"
  /protein_id="AAF97184.1"
  /db_xref="GI:9664581"
  /translation="MNECAVCGILYLAGSPCAGCQSLRAQOQSFSDSDAALPTEVPG
  LIDAAQWYDLEGIKAPVEEAAASLPFGGSESTNISRLEPFGISHPEGIIPFEL
  SNQDSQEEEVQHVPEVIESTNLEPVVALETEQDSPPVVSASVEIEPVLAPVAPD
  LPLIEBPVQVPPVPGPIRLTAIPFVEPTVDTDLNENWISIESFGGSDQIIFTAEDVIE
  VVFSLEDIVHVEHDGENYPOWEKDSVYSEKSLQFPLHPAQALMVDVGTDSARAL
  LESGPTSIGGQWRSAASFOQIVARPDQSGAMNNGYGTALLQVAVMWSNDPIEVS
  NSATQPEAAILSRVNRINSQSPESINLSQALLISGEEKALLDMAGNESQRES
  NFNTRAAILAQGYDEARLVLVQADKLASGNLLKLAL"
  complement(10248..11969)
  /note="37F11#8"
  /codon_start=1
  /evidence=not experimental
  /transl_table=11
  /product="permease"
  /protein_id="AAF97184.1"
  /db_xref="GI:9664581"
  /translation="MNECAVCGILYLAGSPCAGCQSLRAQOQSFSDSDAALPTEVPG
  LIDAAQWYDLEGIKAPVEEAAASLPFGGSESTNISRLEPFGISHPEGIIPFEL
  SNQDSQEEEVQHVPEVIESTNLEPVVALETEQDSPPVVSASVEIEPVLAPVAPD
  LPLIEBPVQVPPVPGPIRLTAIPFVEPTVDTDLNENWISIESFGGSDQIIFTAEDVIE
  VVFSLEDIVHVEHDGENYPOWEKDSVYSEKSLQFPLHPAQALMVDVGTDSARAL
  LESGPTSIGGQWRSAASFOQIVARPDQSGAMNNGYGTALLQVAVMWSNDPIEVS
  NSATQPEAAILSRVNRINSQSPESINLSQALLISGEEKALLDMAGNESQRES
  NFNTRAAILAQGYDEARLVLVQADKLASGNLLKLAL"
  complement(10248..11969)
  /note="37F11#8"
  /codon_start=1
  /evidence=not experimental
  /transl_table=11
  /product="permease"
  /protein_id="AAF97184.1"
  /db_xref="GI:9664581"
  /translation="MNECAVCGILYLAGSPCAGCQSLRAQOQSFSDSDAALPTEVPG
  LIDAAQWYDLEGIKAPVEEAAASLPFGGSESTNISRLEPFGISHPEGIIPFEL
  SNQDSQEEEVQHVPEVIESTNLEPVVALETEQDSPPVVSASVEIEPVLAPVAPD
  LPLIEBPVQVPPVPGPIRLTAIPFVEPTVDTDLNENWISIESFGGSDQIIFTAEDVIE
  VVFSLEDIVHVEHDGENYPOWEKDSVYSEKSLQFPLHPAQALMVDVGTDSARAL
  LESGPTSIGGQWRSAASFOQIVARPDQSGAMNNGYGTALLQVAVMWSNDPIEVS
  NSATQPEAAILSRVNRINSQSPESINLSQALLISGEEKALLDMAGNESQRES
  NFNTRAAILAQGYDEARLVLVQADKLASGNLLKLAL"
  complement(10248..11969)
  /note="37F11#8"
  /codon_start=1
  /evidence=not experimental
  /transl_table=11
  /product="permease"
  /protein_id="AAF97184.1"
  /db_xref="GI:9664581"
  /translation="MNECAVCGILYLAGSPCAGCQSLRAQOQSFSDSDAALPTEVPG
  LIDAAQWYDLEGIKAPVEEAAASLPFGGSESTNISRLEPFGISHPEGIIPFEL
  SNQDSQEEEVQHVPEVIESTNLEPVVALETEQDSPPVVSASVEIEPVLAPVAPD
  LPLIEBPVQVPPVPGPIRLTAIPFVEPTVDTDLNENWISIESFGGSDQIIFTAEDVIE
  VVFSLEDIVHVEHDGENYPOWEKDSVYSEKSLQFPLHPAQALMVDVGTDSARAL
  LESGPTSIGGQWRSAASFOQIVARPDQSGAMNNGYGTALLQVAVMWSNDPIEVS
  NSATQPEAAILSRVNRINSQSPESINLSQALLISGEEKALLDMAGNESQRES
  NFNTRAAILAQGYDEARLVLVQADKLASGNLLKLAL"
  complement(10248..11969)
  /note="37F11#8"
  /codon_start=1
  /evidence=not experimental
  /transl_table=11
  /product="permease"
  /protein_id="AAF97184.1"
  /db_xref="GI:9664581"
  /translation="MNECAVCGILYLAGSPCAGCQSLRAQOQSFSDSDAALPTEVPG
  LIDAAQWYDLEGIKAPVEEAAASLPFGGSESTNISRLEPFGISHPEGIIPFEL
  SNQDSQEEEVQHVPEVIESTNLEPVVALETEQDSPPVVSASVEIEPVLAPVAPD
  LPLIEBPVQVPPVPGPIRLTAIPFVEPTVDTDLNENWISIESFGGSDQIIFTAEDVIE
  VVFSLEDIVHVEHDGENYPOWEKDSVYSEKSLQFPLHPAQALMVDVGTDSARAL
  LESGPTSIGGQWRSAASFOQIVARPDQSGAMNNGYGTALLQVAVMWSNDPIEVS
  NSATQPEAAILSRVNRINSQSPESINLSQALLISGEEKALLDMAGNESQRES
  NFNTRAAILAQGYDEARLVLVQADKLASGNLLKLAL"
  complement(10248..11969)
  /note="37F11#8"
  /codon_start=1
  /evidence=not experimental
  /transl_table=11
  /product="permease"
  /protein_id="AAF97184.1"
  /db_xref="GI:9664581"
  /translation="MNECAVCGILYLAGSPCAGCQSLRAQOQSFSDSDAALPTEVPG
  LIDAAQWYDLEGIKAPVEEAAASLPFGGSESTNISRLEPFGISHPEGIIPFEL
  SNQDSQEEEVQHVPEVIESTNLEPVVALETEQDSPPVVSASVEIEPVLAPVAPD
  LPLIEBPVQVPPVPGPIRLTAIPFVEPTVDTDLNENWISIESFGGSDQIIFTAEDVIE
  VVFSLEDIVHVEHDGENYPOWEKDSVYSEKSLQFPLHPAQALMVDVGTDSARAL
  LESGPTSIGGQWRSAASFOQIVARPDQSGAMNNGYGTALLQVAVMWSNDPIEVS
  NSATQPEAAILSRVNRINSQSPESINLSQALLISGEEKALLDMAGNESQRES
  NFNTRAAILAQGYDEARLVLVQADKLASGNLLKLAL"
  complement(10248..11969)
  /note="37F11#8"
  /codon_start=1
  /evidence=not experimental
  /transl_table=11
  /product="permease"
  /protein_id="AAF97184.1"
  /db_xref="GI:9664581"
  /translation="MNECAVCGILYLAGSPCAGCQSLRAQOQSFSDSDAALPTEVPG
  LIDAAQWYDLEGIKAPVEEAAASLPFGGSESTNISRLEPFGISHPEGIIPFEL
  SNQDSQEEEVQHVPEVIESTNLEPVVALETEQDSPPVVSASVEIEPVLAPVAPD
  LPLIEBPVQVPPVPGPIRLTAIPFVEPTVDTDLNENWISIESFGGSDQIIFTAEDVIE
  VVFSLEDIVHVEHDGENYPOWEKDSVYSEKSLQFPLHPAQALMVDVGTDSARAL
  LESGPTSIGGQWRSAASFOQIVARPDQSGAMNNGYGTALLQVAVMWSNDPIEVS
  NSATQPEAAILSRVNRINSQSPESINLSQALLISGEEKALLDMAGNESQRES
  NFNTRAAILAQGYDEARLVLVQADKLASGNLLKLAL"
  complement(10248..11969)
  /note="37F11#8"
  /codon_start=1
  /evidence=not experimental
  /transl_table=11
  /product="permease"
  /protein_id="AAF97184.1"
  /db_xref="GI:9664581"
  /translation="MNECAVCGILYLAGSPCAGCQSLRAQOQSFSDSDAALPTEVPG
  LIDAAQWYDLEGIKAPVEEAAASLPFGGSESTNISRLEPFGISHPEGIIPFEL
  SNQDSQEEEVQHVPEVIESTNLEPVVALETEQDSPPVVSASVEIEPVLAPVAPD
  LPLIEBPVQVPPVPGPIRLTAIPFVEPTVDTDLNENWISIESFGGSDQIIFTAEDVIE
  VVFSLEDIVHVEHDGENYPOWEKDSVYSEKSLQFPLHPAQALMVDVGTDSARAL
  LESGPTSIGGQWRSAASFOQIVARPDQSGAMNNGYGTALLQVAVMWSNDPIEVS
  NSATQPEAAILSRVNRINSQSPESINLSQALLISGEEKALLDMAGNESQRES
  NFNTRAAILAQGYDEARLVLVQADKLASGNLLKLAL"
  complement(10248..11969)
  /note="37F11#8"
  /codon_start=1
  /evidence=not experimental
  /transl_table=11
  /product="permease"
  /protein_id="AAF97184.1"
  /db_xref="GI:9664581"
  /translation="MNECAVCGILYLAGSPCAGCQSLRAQOQSFSDSDAALPTEVPG
  LIDAAQWYDLEGIKAPVEEAAASLPFGGSESTNISRLEPFGISHPEGIIPFEL
  SNQDSQEEEVQHVPEVIESTNLEPVVALETEQDSPPVVSASVEIEPVLAPVAPD
  LPLIEBPVQVPPVPGPIRLTAIPFVEPTVDTDLNENWISIESFGGSDQIIFTAEDVIE
  VVFSLEDIVHVEHDGENYPOWEKDSVYSEKSLQFPLHPAQALMVDVGTDSARAL
  LESGPTSIGGQWRSAASFOQIVARPDQSGAMNNGYGTALLQVAVMWSNDPIEVS
  NSATQPEAAILSRVNRINSQSPESINLSQALLISGEEKALLDMAGNESQRES
  NFNTRAAILAQGYDEARLVLVQADKLASGNLLKLAL"
  complement(10248..11969)
  /note="37F11#8"
  /codon_start=1
  /evidence=not experimental
  /transl_table=11
  /product="permease"
  /protein_id="AAF97184.1"
  /db_xref="GI:9664581"
  /translation="MNECAVCGILYLAGSPCAGCQSLRAQOQSFSDSDAALPTEVPG
  LIDAAQWYDLEGIKAPVEEAAASLPFGGSESTNISRLEPFGISHPEGIIPFEL
  SNQDSQEEEVQHVPEVIESTNLEPVVALETEQDSPPVVSASVEIEPVLAPVAPD
  LPLIEBPVQVPPVPGPIRLTAIPFVEPTVDTDLNENWISIESFGGSDQIIFTAEDVIE
  VVFSLEDIVHVEHDGENYPOWEKDSVYSEKSLQFPLHPAQALMVDVGTDSARAL
  LESGPTSIGGQWRSAASFOQIVARPDQSGAMNNGYGTALLQVAVMWSNDPIEVS
  NSATQPEAAILSRVNRINSQSPESINLSQALLISGEEKALLDMAGNESQRES
  NFNTRAAILAQGYDEARLVLVQADKLASGNLLKLAL"
  complement(10248..11969)
  /note="37F11#8"
  /codon_start=1
  /evidence=not experimental
  /transl_table=11
  /product="permease"
  /protein_id="AAF97184.1"
  /db_xref="GI:9664581"
  /translation="MNECAVCGILYLAGSPCAGCQSLRAQOQSFSDSDAALPTEVPG
  LIDAAQWYDLEGIKAPVEEAAASLPFGGSESTNISRLEPFGISHPEGIIPFEL
  SNQDSQEEEVQHVPEVIESTNLEPVVALETEQDSPPVVSASVEIEPVLAPVAPD
  LPLIEBPVQVPPVPGPIRLTAIPFVEPTVDTDLNENWISIESFGGSDQIIFTAEDVIE
  VVFSLEDIVHVEHDGENYPOWEKDSVYSEKSLQFPLHPAQALMVDVGTDSARAL
  LESGPTSIGGQWRSAASFOQIVARPDQSGAMNNGYGTALLQVAVMWSNDPIEVS
  NSATQPEAAILSRVNRINSQSPESINLSQALLISGEEKALLDMAGNESQRES
  NFNTRAAILAQGYDEARLVLVQADKLASGNLLKLAL"
  complement(10248..11969)
  /note="37F11#8"
  /codon_start=1
  /evidence=not experimental
  /transl_table=11
  /product="permease"
  /protein_id="AAF97184.1"
  /db_xref="GI:9664581"
  /translation="MNECAVCGILYLAGSPCAGCQSLRAQOQSFSDSDAALPTEVPG
  LIDAAQWYDLEGIKAPVEEAAASLPFGGSESTNISRLEPFGISHPEGIIPFEL
  SNQDSQEEEVQHVPEVIESTNLEPVVALETEQDSPPVVSASVEIEPVLAPVAPD
  LPLIEBPVQVPPVPGPIRLTAIPFVEPTVDTDLNENWISIESFGGSDQIIFTAEDVIE
  VVFSLEDIVHVEHDGENYPOWEKDSVYSEKSLQFPLHPAQALMVDVGTDSARAL
  LESGPTSIGGQWRSAASFOQIVARPDQSGAMNNGYGTALLQVAVMWSNDPIEVS
  NSATQPEAAILSRVNRINSQSPESINLSQALLISGEEKALLDMAGNESQRES
  NFNTRAAILAQGYDEARLVLVQADKLASGNLLKLAL"
  complement(10248..11969)
  /note="37F11#8"
  /codon_start=1
  /evidence=not experimental
  /transl_table=11
  /product="permease"
  /protein_id="AAF97184.1"
  /db_xref="GI:9664581"
  /translation="MNECAVCGILYLAGSPCAGCQSLRAQOQSFSDSDAALPTEVPG
  LIDAAQWYDLEGIKAPVEEAAASLPFGGSESTNISRLEPFGISHPEGIIPFEL
  SNQDSQEEEVQHVPEVIESTNLEPVVALETEQDSPPVVSASVEIEPVLAPVAPD
  LPLIEBPVQVPPVPGPIRLTAIPFVEPTVDTDLNENWISIESFGGSDQIIFTAEDVIE
  VVFSLEDIVHVEHDGENYPOWEKDSVYSEKSLQFPLHPAQALMVDVGTDSARAL
  LESGPTSIGGQWRSAASFOQIVARPDQSGAMNNGYGTALLQVAVMWSNDPIEVS
  NSATQPEAAILSRVNRINSQSPESINLSQALLISGEEKALLDMAGNESQRES
  NFNTRAAILAQGYDEARLVLVQADKLASGNLLKLAL"
  complement(10248..11969)
  /note="37F11#8"
  /codon_start=1
  /evidence=not experimental
  /transl_table=11
  /product="permease"
  /protein_id="AAF97184.1"
  /db_xref="GI:9664581"
  /translation="MNECAVCGILYLAGSPCAGCQSLRAQOQSFSDSDAALPTEVPG
  LIDAAQWYDLEGIKAPVEEAAASLPFGGSESTNISRLEPFGISHPEGIIPFEL
  SNQDSQEEEVQHVPEVIESTNLEPVVALETEQDSPPVVSASVEIEPVLAPVAPD
  LPLIEBPVQVPPVPGPIRLTAIPFVEPTVDTDLNENWISIESFGGSDQIIFTAEDVIE
  VVFSLEDIVHVEHDGENYPOWEKDSVYSEKSLQFPLHPAQALMVDVGTDSARAL
  LESGPTSIGGQWRSAASFOQIVARPDQSGAMNNGYGTALLQVAVMWSNDPIEVS
  NSATQPEAAILSRVNRINSQSPESINLSQALLISGEEKALLDMAGNESQRES
  NFNTRAAILAQGYDEARLVLVQADKLASGNLLKLAL"
  complement(10248..11969)
  /note="37F11#8"
  /codon_start=1
  /evidence=not experimental
  /transl_table=11
  /product="permease"
  /protein_id="AAF97184.1"
  /db_xref="GI:9664581"
  /translation="MNECAVCGILYLAGSPCAGCQSLRAQOQSFSDSDAALPTEVPG
  LIDAAQWYDLEGIKAPVEEAAASLPFGGSESTNISRLEPFGISHPEGIIPFEL
  SNQDSQEEEVQHVPEVIESTNLEPVVALETEQDSPPVVSASVEIEPVLAPVAPD
  LPLIEBPVQVPPVPGPIRLTAIPFVEPTVDTDLNENWISIESFGGSDQIIFTAEDVIE
  VVFSLEDIVHVEHDGENYPOWEKDSVYSEKSLQFPLHPAQALMVDVGTDSARAL
  LESGPTSIGGQWRSAASFOQIVARPDQSGAMNNGYGTALLQVAVMWSNDPIEVS
  NSATQPEAAILSRVNRINSQSPESINLSQALLISGEEKALLDMAGNESQRES
  NFNTRAAILAQGYDEARLVLVQADKLASGNLLKLAL"
  complement(10248..11969)
  /note="37F11#8"
  /codon_start=1
  /evidence=not experimental
  /transl_table=11
  /product="permease"
  /protein_id="AAF97184.1"
  /db_xref="GI:9664581"
  /translation="MNECAVCGILYLAGSPCAGCQSLRAQOQSFSDSDAALPTEVPG
  LIDAAQWYDLEGIKAPVEEAAASLPFGGSESTNISRLEPFGISHPEGIIPFEL
  SNQDSQEEEVQHVPEVIESTNLEPVVALETEQDSPPVVSASVEIEPVLAPVAPD
  LPLIEBPVQVPPVPGPIRLTAIPFVEPTVDTDLNENWISIESFGGSDQIIFTAEDVIE
  VVFSLEDIVHVEHDGENYPOWEKDSVYSEKSLQFPLHPAQALMVDVGTDSARAL
  LESGPTSIGGQWRSAASFOQIVARPDQSGAMNNGYGTALLQVAVMWSNDPIEVS
  NSATQPEAAILSRVNRINSQSPESINLSQALLISGEEKALLDMAGNESQRES
  NFNTRAAILAQGYDEARLVLVQADKLASGNLLKLAL"
  complement(10248..11969)
  /note="37F11#8"
  /codon_start=1
  /evidence=not experimental
  /transl_table=11
  /product="permease"
  /protein_id="AAF97184.1"
  /db_xref="GI:9664581"
  /translation="MNECAVCGILYLAGSPCAGCQSLRAQOQSFSDSDAALPTEVPG
  LIDAAQWYDLEGIKAPVEEAAASLPFGGSESTNISRLEPFGISHPEGIIPFEL
  SNQDSQEEEVQHVPEVIESTNLEPVVALETEQDSPPVVSASVEIEPVLAPVAPD
  LPLIEBPVQVPPVPGPIRLTAIPFVEPTVDTDLNENWISIESFGGSDQIIFTAEDVIE
  VVFSLEDIVHVEHDGENYPOWEKDSVYSEKSLQFPLHPAQALMVDVGTDSARAL
  LESGPTSIGGQWRSAASFOQIVARPDQSGAMNNGYGTALLQVAVMWSNDPIEVS
  NSATQPEAAILSRVNRINSQSPESINLSQALLISGEEKALLDMAGNESQRES
  NFNTRAAILAQGYDEARLVLVQADKLASGNLLKLAL"
  complement(10248..11969)
  /note="37F11#8"
  /codon_start=1
  /evidence=not experimental
  /transl_table=11
  /product="permease"
  /protein_id="AAF97184.1"
  /db_xref="GI:9664581"
  /translation="MNECAVCGILYLAGSPCAGCQSLRAQOQSFSDSDAALPTEVPG
  LIDAAQWYDLEGIKAPVEEAAASLPFGGSESTNISRLEPFGISHPEGIIPFEL
  SNQDSQEEEVQHVPEVIESTNLEPVVALETEQDSPPVVSASVEIEPVLAPVAPD
  LPLIEBPVQVPPVPGPIRLTAIPFVEPTVDTDLNENWISIESFGGSDQIIFTAEDVIE
  VVFSLEDIVHVEHDGENYPOWEKDSVYSEKSLQFPLHPAQALMVDVGTDSARAL
  LESGPTSIGGQWRSAASFOQIVARPDQSGAMNNGYGTALLQVAVMWSNDPIEVS
  NSATQPEAAILSRVNRINSQSPES
```

```
/db_xref="GI:9664581"
/translacion="MGNILFSKRHPAQLIALISLFLVGTWQFIRSHAAAGGDDIDL
TVTRDGFENQAVYVGLGILLFVYALIIIEVHRTLAATGLGLLAVIALAHYSVEEAL
SUKAVTIDWTTIGLLGMVMVGIIISHTGVFEVAVQAYKKGSGEITWTLVLICSV
TAVLSAFIDNWTMLLPVTIKIAQVLDKPIPIILISELVFSNIGGAATWIGDPNNI
MIGSAMDPAISGNKDPNIANLKSQVTFNDPIETEMAGILMTIVPSFMLIRMLYKDE
FSGRDRDTELEDRQTAIKDKYKSGAILTLVLQGFELFVHILAVSWALSGALV
MLLANTHELEPLEHEVETWTLFFAGLFLVHLSQYMGVVDIYGEYVERKAIFPPD
YRLAAATILWVASIAPADINIPYTATMIPVLSLAFELNPLNPLIMLAFAGCL
GNGTLIGASNVVAGSEEGEYIPISFNEFRAGFPVMLLSTFIVSFYMLIVYVGG
EDGALTWKVLVGLISLLGIIAQYSRGRKAKSPAEALVDDDPDEIVSTIKSVIKTGS
KGSSEEE"
complement(12019..12282)
/notes="37f11#9; contains KH domain"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="RNA-binding protein"
/protein_id="AAF97186.1"
/db_xref="GI:9664583"
/translacion="MSVPNHIIROAKDTSPLITMRIGKSGLTDSVVVELEGOLSSRL
VKVNRGLYPRDELKTVMNMAEKINSSVFAAGNVAFPRN"
complement(12279..12614)
/notes="37f11#10; contains Zn-finger; conserved protein"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="unknown"
/protein_id="AAF97185.1"
/db_xref="GI:9664582"
/translacion="MGRRLNRDKKGHAIWSQTLQSLLNKPNQFTPEQSRFALH
LVKVSRRHRIREFSEVREIMCRNTLLRYGENATIRFNHGIQCHTSVRRIPY
RHSQKVKA"
complement(12624..13274)
/notes="37f11#11"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="adenylate kinase"
/protein_id="AAF97187.1"
/db_xref="GI:9664584"
/translacion="MDEGRQGLNLTLPFGPGAGKGTQOQRMDATGLPOVSTGDM
RAVAGTSVIGIAKSYMKGALVPDSVILDLRLDHDDAKNGVDFGFPRTVPOA
EALSEIVEAVLAIDVPDRIVERICGRYSGCTGTVHDFNPNQNGICGSGFTE
KRADDVEETVAVLAIDVPDRIVERICGRYSGCTGTVHDFNPNQNGICGSGFTE
complement(13284..16691)
/notes="37f11#12"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="membrane-associated subtilysin-type serine
protease"
/protein_id="AAF97188.1"
/db_xref="GI:9664585"
/translacion="MYQLSGTQATVAGGSASLAREVIREVAGINKPSASLIKATL
INGEDLTGPDIPNANGWQIDLENSLPSSGSLDVPDQDERLQAGFSLIYSFD
LDGSGDIDTLAREASANAQSSRLANNLDLILAPDGSVLYNDFFSGISFTG
GSADLNMIERIRIPAGATTQNGDMVTVEHSGSQRYSIIVIAADATLIPKADLTTF
GNSLPSSSPLGLVLSLAWHQTLASGPIRIQFEDITSGLIYSNSSLSEGG
SLDSVSFTPTSTGVHKLSDLTNNQVPELDAINGDNIIDIDITAGGLRVI
FQNDGSIPTQDRDSASVTMDVRNETGLSLFVIAHEGTERGERVSLTSVQBPD
PTYTLTLLSDSKSNQVOTGVFTISGQGEENDTIVLTNLDDTSASFDGATKYAR
AGSFIVDYARYONCPVSHQRYIIEIGOVSDVVVPSGTSGVSAKPGERTGFGYR
PNTGNSAQVMTSCTSSQWIMLGNSSSLFEPLNLQDLSDMVNVFIPSIANGE
PLAGSTQVCTVTSPTSDSLNRVETVDVAVSELKAFPSLDYGPDGVGPGALALPVF
ADTGELVFHNTKNGVPLDFAVTLBERNGWAAEQDEQISTSLSVTLGPQOS
GDVEMRLVLPETAREGNTYLRVSESSQPFNTLSLVVGENGLVNLVSQVGLTIA
APVNDRTFTEFIVENSGNDLDLEWSTQPDWGSIGSNPPTSPVLSQASVQAL
KAPNOTASGFDQMLFVNGTNRFTNAELLVIEVATSPAGISVSDTAPLFI
PRGDSQKQITFMQNLPLSGQTVREDBEGLNLTDRSSTISPEBITDLAIGSIE
VVGKVSIDEDSIDGEMNLVLTITDGVWIEFWADTSVSSQSSGGIFGLPAYLVYP
LVLLALGLYGRRLKQSKMDDDGTIELVAPDAHTADHLGTRERQALDISHVNDI
ASGEVSDQELAAALAQSLDMPFVTKAQVPTGTRPPSGLPNSGLFPVGLPPAGL
```

```
CDS
PPAKSVPLVLPQVYTAGPPLPPGGLPDGWTMEQNHVGEQYLQRMGLN"
complement(16790..18130)
/notes="37f11#13"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="membrane-associated subtilysin-type serine
protease"
/protein_id="AAF97189.1"
/db_xref="GI:9664586"

Alignment Scores:
Pred. No.: 1,02e-29 Length: 60006
Score: 626.50 Matches: 165
Percent Similarity: 53.76% Conservative: 78
Best Local Similarity: 36.50% Mismatches: 150
Query Match: 27.91% Indels: 60
DB: 1 Gaps: 17

US-09-985-689A-2 (1-434) x AF268611 (1-60006)
Qy 13 AlaglnSerSerTyrglyLeuTyrglyGlnGlyGlnleValAlaValAlaAaspThrGly 32
Db 17383 GCTCCTCTCTCTTTGAATCTTGATGTTCTGGAGAACGATTCAGTGTGATCGGC 17324
Qy 33 LeuAepThrGlyArgAenAaspSerSerMetHisGluAlaPheArgGlyLysIleThrAla 52
Db 17323 GTCGATATGGAT-----CATCCTGACCTCATCGACGGGTTCAGCA 17282
Qy 53 LeuTyrala---LeuGlyArgThrAsnAsnAlaAaspThrAsn---GlyHisGlyThr 70
Db 17281 ATTAATCTCAGTTCGGTTTAGATGTTCTCCAAATGATAGCAACAGTGTGATCGGACG 17222
Qy 71 HisValAlaGlySerValLeuGlyAsnGly-----AlathrAsnLysGlyMetAla 87
Db 17221 CACGTGGCTAGCTGCTTGGCGATGTCAGCGTGATGACAGTAGCAGCGGTGTGCT 17162
Qy 88 ProGlnAlaAsnLeuValPheGlnSerIleMetAaspSerSerGlyGlyLeuGlyGlyLeu 107
Db 17161 CCAAGGCAATCTTGATGTCGCTCTCTGATGCCGAATATAAACAGCAAGTGTGCGAGTC 17102
Qy 108 ProSerAenLeuGlnThrLeuPheSerGlnAlaPheSerAlaGlyAlaArgIleHisThr 127
Db 17101 CAAGTTCATCTATGATCTCTCTGATGCCGAATATAAACAGCAAGTGTGCGAGTC 17042
Qy 128 AenSerTrpGlyAlaAlaValAsn---GlyAlaTyThrThrAaspSerArgAenValAasp 146
Db 17041 AATGATGGGTCTTAATGGAATTTACGAGCCTTACACTGCTGATCA-AGATCCACAGAT 16983
Qy 147 AspTyrValAlaGlyAsnAaspMetThrIle---LeuPheAlaAlaGlyAsnGluGlyPro 165
Db 16982 CAATATGCTCCACNATAAATATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 16923
Qy 166 AenGlyGlyThr---IleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAla 184
Db 16922 ACAGGTTCTTCAAAGATTACTGCTCCAGGAACACCCCAAGATGTTCTTTCATTGGCTCA 16863
Qy 185 ThrGluAenLeuAargProSerPheGlySerTyralaAaspAenIleAenHisValAlaGln 204
Db 16862 ACGGTATAT-----GGTAGC-----GTGAGCTCC 16839
Qy 205 PheSerSerArgGlyProThrLysAaspGlyArgIleLysProAaspValMetAlaProGly 224
Db 16838 TATTCATCGAAGGTCCAACTCTCGATGGAAGAATTAACCTGATTTAGTCGCTCCCGGT 16779
Qy 225 ThrTyrlleLeuSerAlaAargSerSerLeuAlaProAasp----- 237
Db 16778 GACGGTATCTGTTCTGACGAGCAGAGAACAAAGACGTTGTTGTTCACTCTGCGGT 16719
Qy 238 SerSerPheTrpAlaAsnHisAaspSerLysTyralaTyMetGlyGlyThrSerMetAla 257
Db 16718 ATTGGTACTCACTCAATGGTTCGTTCAATGTATATGCAACTCAGCGGAACCTTCTCAAGCA 16659
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 3, 2004, 16:37:54 ; Search time 336.684 Seconds
(without alignments)
5476.111 Million cell updates/sec

Title: US-09-985-689A-2
Perfect score: 2245
Sequence: 1 NDVARGIVKADVAQSSVGLY.....EVOAYNVPGPQNFSLAIVN 434

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q/cgn2 1/USPTO spool/US09985689/runat 31032004 161806 4145/app query.fasta_1.3498
-DB=N Geneseq 29Jan04 -QFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blomsum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09985689@cgn 1 1 1238/runat 31032004 161806 4145 -NCFU=6 -ICPU=3
-NO_WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq 29Jan04:*

1: Geneseqn1980s:*
2: Geneseqn1990s:*
3: Geneseqn2000s:*
4: Geneseqn2001as:*
5: Geneseqn2001bs:*
6: Geneseqn2002as:*
7: Geneseqn2003as:*
8: Geneseqn2003bs:*
9: Geneseqn2003cs:*
10: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|----|--------------------|
| 1 | 2237 | 99.6 | 1920 | 2 | AAx37277 Bacillus |
| 2 | 2197 | 97.9 | 1923 | 2 | AAx37279 Bacillus |
| 3 | 2191 | 97.6 | 1923 | 2 | AAx37278 Bacillus |
| 4 | 2118.5 | 94.4 | 3003 | 2 | AAV82382 Bacillus |
| 5 | 1998.5 | 89.0 | 1299 | 2 | AAQ27516 Alkali-pr |
| 6 | 457.5 | 20.4 | 1977 | 2 | AAx85667 Thermococ |
| 7 | 457.5 | 20.4 | 1977 | 2 | AAx05926 WO9856926 |
| 8 | 425.5 | 19.0 | 1236 | 2 | AAx05920 Hyperther |

| | | | | | |
|----|-------|------|--------|---|--------------------------------|
| 9 | 425.5 | 19.0 | 1566 | 2 | AAx85668 Pyrococu |
| 10 | 425.5 | 19.0 | 1962 | 2 | AAx85695 Pyrococu |
| 11 | 425.5 | 19.0 | 1862 | 2 | AAx05929 Hyperther |
| 12 | 414.5 | 18.5 | 1977 | 2 | AAx85669 Pyrococu |
| 13 | 383 | 17.1 | 2121 | 4 | ABL54900 T. yonsei |
| 14 | 354.5 | 15.8 | 2539 | 2 | AAx85669 Pyrococu |
| 15 | 354.5 | 15.8 | 2809 | 2 | AAx85669 Pyrococu |
| 16 | 343 | 15.3 | 135638 | 7 | ABx34289 S. atrool |
| 17 | 338.5 | 15.1 | 1329 | 6 | ABx74643 Bacillus |
| 18 | 316.5 | 14.1 | 3413 | 2 | AAx72330 F. balust |
| 19 | 311.5 | 13.9 | 1859 | 2 | AAx85677 Thermococ |
| 20 | 308.5 | 13.7 | 1560 | 7 | ABx37569 Streptomy |
| 21 | 308.5 | 13.7 | 59816 | 7 | ABx37516 Streptomy |
| 22 | 308.5 | 13.7 | 59816 | 7 | ABx37515 Streptomy |
| 23 | 307 | 13.7 | 2835 | 2 | AAx08141 Hyperther |
| 24 | 307 | 13.7 | 4765 | 2 | AAx08132 Pyrococu |
| 25 | 307 | 13.7 | 4765 | 2 | AAx85670 Pyrococu |
| 26 | 307 | 13.7 | 4765 | 2 | AAx05921 WO9856926 |
| 27 | 302.5 | 13.5 | 2532 | 2 | AAQ29134 Encodes R |
| 28 | 300 | 13.4 | 898 | 2 | AAx08131 Hyperther |
| 29 | 291.5 | 13.0 | 564 | 2 | AAx08134 DNA seque |
| 30 | 291.5 | 13.0 | 564 | 2 | AAx85676 Thermococ |
| 31 | 283 | 12.6 | 3788 | 9 | ADD24905 DNA encod |
| 32 | 282 | 12.6 | 546 | 4 | ABL53453 T. yonsei |
| 33 | 277 | 12.3 | 3743 | 3 | ADD24901 DNA encod |
| 34 | 273 | 12.2 | 1306 | 6 | ABL55784 Bacillus |
| 35 | 273 | 12.2 | 1330 | 6 | ABL55787 Bacillus |
| 36 | 269.5 | 12.0 | 4716 | 9 | ABQ80437 CspA codi |
| 37 | 269.5 | 12.0 | 4740 | 7 | ABx258957 Group B S |
| 38 | 262.5 | 11.7 | 2273 | 2 | AAQ04339 Aquaticin |
| 39 | 262.5 | 11.7 | 2274 | 2 | AAQ12838 Aquaticin |
| 40 | 262.5 | 11.7 | 2274 | 2 | AAQ75859 Aqualytin |
| 41 | 262.5 | 11.7 | 4850 | 6 | ABN71526 Streptoco |
| 42 | 262.5 | 11.7 | 4710 | 6 | ABN71162 Streptoco |
| 43 | 262.5 | 11.7 | 4734 | 6 | ABN70525 Streptoco |
| 44 | 262.5 | 11.7 | 4770 | 6 | ABN69191 Streptoco |
| 45 | 262.5 | 11.7 | 110000 | 6 | ABN71527_19 Continuation (20 o |

ALIGNMENTS

RESULT 1

AAx37277

ID AAx37277 standard; DNA; 1920 BP.

XX AC AAx37277;

XX AC AAx37277;

XX DT 20-MAR-2003 (revised)

XX DT 21-JUL-1999 (first entry)

XX DE Bacillus alkaline protease encoding DNA.

XX KW Alkaline protease; Bacillus; casein digestion; oleic acid; enzyme;

XX KW washing composition; oxidising agent; ss.

XX OS Bacillus sp.

XX PN WO9918218-A1.

XX PD 15-APR-1999.

XX PF 07-OCT-1998; 98WO-JP004528.

XX PR 07-OCT-1997; 97JP-00274570.

XX PA (KAOS) KAO CORP.

XX PI Takaiwa M, Okuda M, Saeki K, Kubota H, Hitomi J, Kageyama Y;

XX PI Shikata S, Nomura M;

XX DR WPI; 1999-2877736/27.

XX DR P-PSDB; AAY17087, AAY17089.

PT Alkali protease from Bacillus used in washing powders.

XX Disclosure; Page 53-58; 71pp; Japanese.

XX The invention relates to alkaline proteases produced by strains of
CC Bacillus. The proteases ability to digest casein is not inhibited by
CC oleic acid and they have a high stability to oxidising agents. The
CC alkaline protease of the invention has the following properties: (a) it
CC is active over the pH range 4-13 and has at least 80% of its optimum
CC activity over the pH range 6-12; (b) after 30 minutes at 40 deg. C it is
CC stable over the pH range 6-11; (c) its isoelectric point is 8.9-9.1; (d)
CC its ability to digest casein is not inhibited by oleic acid; (e) it has
CC molecular weight about 43,000 by SDS-PAGE. The alkaline proteases can be
CC used as enzymes in washing compositions for use in automatic dishwashers
CC and for washing clothes. The stability to oxidising agents allows the
CC enzyme to be an effective component of washing compositions including
CC bleaches. The present sequence represents an alkaline protease encoding
CC DNA. (Updated on 20-MAR-2003 to correct DR field.)

XX Sequence 1920 BP; 576 A; 411 C; 473 G; 460 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 6,516-169 Length: 1920
Score: 2237.00 Matches: 433
Percent Similarity: 99.77% Conservative: 0
Best Local Similarity: 99.77% Mismatches: 1
Query Match: 99.64% Indels: 0
DB: 2 Gaps: 0

US-09-985-689A-2 (1-434) x AAX37277 (1-1920)

| | | | |
|----|------|---|------|
| QY | 1 | AsnAspValAlaArgGlyLeuValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyr | 20 |
| DB | 616 | AATGATGTGCCAGAGGTATTGTCAAGCGGATGTGGCACAGACAGCTACGGTTGTAT | 675 |
| QY | 21 | GlyGlnGlyGlnLeuValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer | 40 |
| DB | 676 | GGACAGGCCAGATTGTCCGAGTTGCCGATCTGGATTGGATACAGGAAGAACACACAGT | 735 |
| QY | 41 | SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn | 60 |
| DB | 736 | TCGATGCATGAAGCCITCCGCGGTAAATAACAGCACTATATGCACTGGGTGGCGAAT | 795 |
| QY | 61 | AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly | 80 |
| DB | 796 | AATGCGAATGATACGACGCTCATGTGTACCCATGTCGAGGTTCCGATTAGAAATGCG | 855 |
| QY | 81 | AlaThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer | 100 |
| DB | 856 | GCAACGAATAAGGAATGGACCTCAAGCGAATCTGGTTTTCAATCCATCATGGATAGC | 915 |
| QY | 101 | SerGlyGlyLeuGlyLeuProSerAsnLeuGlnThrIleuPheSerGlnAlaPheSer | 120 |
| DB | 916 | AGTGGTGGGCTTGGAGCTTGCCTTCCATCTCGAACACCTTATTCAGCCAAAGCAITTCAGT | 975 |
| QY | 121 | AlaGlyAlaArgIleHisThrAsnSerTyrGlyValAlaAlaValAsnGlyAlaTyrThrThr | 140 |
| DB | 976 | GCAGTGTCCAGAAATTCATACAACTCTCGGGGCGACGCGTGAATGGGCGCTACACGACA | 1035 |
| QY | 141 | AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleuPheAlaAla | 160 |
| DB | 1036 | GATTCAGAAATGTGGATGACTATGTAAAGAAATAATGATATGACGATTTCTTCGCGGT | 1095 |
| QY | 161 | GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle | 180 |
| DB | 1096 | GGGNATGAAGCGCGGCGGTACCATCAGTGCACCTGGTACGCTAAGAAACGCCATA | 1155 |
| QY | 181 | ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn | 200 |
| DB | 1156 | ACAGTCGCGCGAACCGAAACCTCGCTCCAGCTTCGTTCCATGCGAGATAATATTAAC | 1215 |
| QY | 201 | HisValAlaGlnPheSerArgGlyProThrLysAspGlyArgIleLysProAspVal | 220 |

| | | | |
|----|------|---|------|
| DB | 1216 | CACGTTGCACAGTTCTCTTCCCGCCGCAAAAGATGGCGAATCAAGCCTGATGTC | 1275 |
| QY | 221 | MetAlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe | 240 |
| DB | 1276 | ATGGCCCGCAGGACATACATTTTATCAGCAAGATCTCTCTTGCACCCGATTCCTCTTC | 1335 |
| QY | 241 | TpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle | 260 |
| DB | 1336 | TGGCGCAATCATGACAGCAAAATATGCTATATGGGTGGAAACGTCATGCGCAACCGAAT | 1395 |
| QY | 261 | ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro | 280 |
| DB | 1396 | GTTGCGGGGAATGTTCCACAGCTCCGTGAGCATTTTGTGAAAATAGAGGAATCACTCT | 1455 |
| QY | 281 | LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspValGlyLeuGlyTyr | 300 |
| DB | 1456 | ARCCCTTCCCTATTGAAGCAGCTTTGATTGAGGTGCTGCTGATGTTGGATTGGTTAT | 1515 |
| QY | 301 | ProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr | 320 |
| DB | 1516 | CCGAACGGAACCAAGGATGGGCGCGAGTGACCCCTGGATAAAATCGTTGAACGTTGCCAT | 1575 |
| QY | 321 | ValAsnGluSerSerAlaLeuSerThrSerGlnLysAlaAlaThrTyrThrPheThrAlaThr | 340 |
| DB | 1576 | GTGAACGATCCAGTGCCTATCACTAGCCAAAAGAGCGACATATACCTTTACTGCAACG | 1635 |
| QY | 341 | AlaGlyLysProLeuLysIleSerLeuValTyrSerAspAlaProAlaSerThrAla | 360 |
| DB | 1636 | CGGGGCAAGCCATTGAAATCTCCCTGTATGTGCGATGCCCTGCCCTGCAAGCACTACTGCT | 1695 |
| QY | 361 | SerValThrLeuValAsnAspLeuValIleThrAlaProAsnGlyThrArgTyr | 380 |
| DB | 1696 | TCGTAAACCTGGTCAATGATTTGGATTGGTTCATTCAGCACCACCAACGNAACAATAT | 1755 |
| QY | 381 | ValGlyAsnAspPheSerAlaProPheAspAsnAsnTyrAspGlyArgAsnAsnValGlu | 400 |
| DB | 1756 | GTGCGGAATGACCTCTCAGCACCATTTGACAAATACTGGGATGGCGCAATAACGTAGAA | 1815 |
| QY | 401 | AsnValPheIleAsnSerProGlnSerGlyTyrThrTyrThrIleGluValGlnAlaTyrAsn | 420 |
| DB | 1816 | AATGATTTTATTAATTCGCCCAAGAGTGGAAACATATACCATTTGAGGTGCAGCATATAT | 1875 |
| QY | 421 | ValProValGlyProGlnAsnPheSerLeuAlaIleValAsn | 434 |
| DB | 1876 | GTGCGGTTGGACCAACAACTTCTCGTTGCGAATTGTGAAC | 1917 |

RESULT 2

AAX37279 standard; DNA; 1923 BP.

XX AAX37279;

XX 20-MAR-2003 (revised)

DT 21-JUL-1999 (first entry)

XX Bacillus alkaline protease encoding DNA.

XX Alkaline protease; Bacillus; casein digestion; oleic acid; enzyme;
KW washing composition; oxidising agent; ss.

XX Bacillus sp.

XX WO9918218-A1.

XX 15-APR-1999.

XX 07-OCT-1998; 98WO-JP004528.

XX 07-OCT-1997; 97JP-00274570.

XX (KAOS) KAO CORP.

XX Takaiwa M, Okuda M, Saeki K, Kubota H, Hitomi J, Kageyama Y;

PI Shikata S, Nomura M;
 XX WPI; 1999-287736/27.
 DR P-PSDB; AAY17091.
 XX
 XX Alkali protease from *Bacillus* used in washing powders.
 PT
 XX
 PS Disclosure; Page 63-68; 71pp; Japanese.
 XX
 CC The invention relates to alkaline proteases produced by strains of
 CC *Bacillus*. The proteases ability to digest casein is not inhibited by
 CC oleic acid and they have a high stability to oxidising agents. The
 CC alkaline protease of the invention has the following properties: (a) it
 CC is active over the pH range 4-13 and has at least 80% of its optimum
 CC activity over the range pH 6-12; (b) after 30 minutes at 40 deg. C it is
 CC stable over the pH range 6-11; (c) its isoelectric point is 8.9-9.1; (d)
 CC its ability to digest casein is not inhibited by oleic acid; (e) it has
 CC molecular weight about 43,000 by SDS-PAGE. The alkaline proteases can be
 CC used as enzymes in washing compositions for use in automatic dishwashers
 CC and for washing clothes. The stability to oxidising agents allows the
 CC enzyme to be an effective component of washing compositions including
 CC bleaches. The present sequence represents an alkaline protease encoding
 CC DNA. (Updated on 20-MAR-2003 to correct DR field.)
 XX
 SQ Sequence 1923 BP; 578 A; 417 C; 474 G; 454 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 1,01e-165 Length: 1923
 Score: 2197.00 Matches: 420
 Percent Similarity: 99.77% Conservative: 13
 Best Local Similarity: 96.77% Mismatches: 1
 Query Match: 97.86% Indels: 0
 DB: 2 Gaps: 0
 US-09-985-689A-2 (1-434) x AAX37279 (1-1923)
 QY 1 AsnAspValAlaArgGlyValValValValAlaAlaGlnSerSerThrGlyLeuTyr 20
 DB 619 AATGATGTTGGCGTGGGAATGTCAAGCGGATGTGGCTCAGACAGCTACGGGTTGTAT 678
 QY 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
 DB 679 GGACAAGACAGATCGTAGCGTTCGCCGATACAGCGTGTGATACAGCTGCGAATGACAGT 738
 QY 41 SerMetHisGluAlaPheArgGlyValValValValAlaLeuTyrAlaLeuGlyArgThrAsn 60
 DB 739 TCGATGTCATGAAGCCCTCCCGGGGAAATTAATCTGCATTATATGCAATGCGGACGACAA 798
 QY 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
 DB 799 AATGCCAATGATACGAATGGTCATGGTACGCAATGGCTGGCTCGTATTAGGAACGGC 858
 QY 81 AlaThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
 DB 859 TCCACTAATAAGGAATAGCGCCTCAGCGCAATAGTCTTCCCAATCTATCATGGATAGC 918
 QY 101 SerGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaPheSer 120
 DB 919 GGTGGGGACTTGGAGGACTACCTTCGAATCTGCAACCTTATTGAGCCAGCATACAGT 978
 QY 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyrThr 140
 DB 979 GCTGGTCCAGAAATTCATACAACTCTCTGGGAGCAGCAGTGAATGGGCTTACACAACA 1038
 QY 141 AspSerArgAsnValAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
 DB 1039 GATTCAGAAATGTGGAGTATGTGCGCAAAATATGATGACGATCCITTTTCGCTGCC 1098
 QY 161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
 DB 1099 GGGATGAGGACCGGACCGCGGACCATCAGTGCACCAGCAGCAGCTAAATGCAATA 1158
 QY 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200

DB 1159 ACAGTCGGAGCTACGGAAAACTCCGCCCAAGCTTTGGTCTTATCGGACAAATATCAAC 1218
 QY 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
 DB 1219 CATGTGGCACATTTCTTTCAGTGGACCGCAAAAGGATGGACGGATCAAAACCGGATGTC 1278
 QY 221 MetAlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerPhe 240
 DB 1279 ATGGCACCGGAAACGTTTATCTATCATCAGCAAGATCTTCTTTGACCGGATTCCTCTTC 1338
 QY 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIle 260
 DB 1339 TGGCGCAACCATGACAGTAAATATGCATACATGCTGGTGGAAACGCTCCATGCTACACCGATC 1398
 QY 261 ValAlaGlyAsnValAlaGlnLeuArgLysHisPheValLysAsnArgGlyIleThrPro 280
 DB 1399 GTTCTTGGAAACGTGGCAGCTTCTGAGCATTCTTGAAAAACAGAGGATCACACA 1458
 QY 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspValGlyLeuGlyTyr 300
 DB 1459 AAGCCTTCTCTATTAAGAGCGCACTGATTGCGGTGCGCTGACATCGGCTTGGCTAC 1518
 QY 301 ProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
 DB 1519 CCGAACGGTACCAAGGATGGGACGAGTGACATTGGATAAATCCCTGAACGTTGCCTAT 1578
 QY 321 ValAsnGluSerSerAlaLeuSerThrSerGlnLysAlaAlaThrTyrThrPheThrAlaThr 340
 DB 1579 GTGACAGAGTCCAGTCTCTATCCACGCAAAAGCGAGCTACTCGTTTACTCTACT 1638
 QY 341 AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrThrAla 360
 DB 1639 GCCGCAAGCCTTTGAAAATCTCCCTGGTATGCTGATGCCCTGCGGACCAACTGCT 1698
 QY 361 SerValThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrArgTyr 380
 DB 1699 TCGTAAACGCTTGTCAATGATCTGGACCTTGTCAATACCGCTCCAAATGGCACACAGTAT 1758
 QY 381 ValGlyAsnAspPheSerAlaProPheAspAsnAsnTrpAspGlyArgAsnValGlu 400
 DB 1759 GTAGGAAATGACTTTTACTTCGCCATACAATGATAAATCGGGATGGCCGCAATACAGTAGAA 1818
 QY 401 AsnValPheIleAsnSerProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
 DB 1819 AATGATTTATTATGACCACAAAGCGGACGTATACAAATTGAAGTACAGGCTTATAC 1878
 QY 421 ValProValGlyProGlnAsnPheSerLeuAlaIleValAsn 434
 DB 1879 GTACCGTTGGACCAAGAACTCTCGTTGGCAATTGTGAAT 1920
 RESULT 3
 AAX37278
 ID AAX37278 standard; DNA; 1923 BP.
 XX
 AC AAX37278;
 XX
 DT 20-MAR-2003 (revised)
 DT 21-JUL-1999 (first entry)
 XX
 DE *Bacillus* alkaline protease encoding DNA.
 XX
 KW Alkaline protease; *Bacillus*; casein digestion; oleic acid; enzyme;
 KW washing composition; oxidising agent; ss.
 XX
 OS *Bacillus* sp.
 XX
 PN WO9918219-A1.
 XX
 PD 15-APR-1999.
 XX
 PF 07-OCT-1998; 98WO-JP004528.
 XX

PR 07-OCT-1997; 97JP-00274570.
 XX (KAOS) KAO CORP.
 XX Takaiwa M, Okuda M, Saeki K, Kubota H, Hitomi J, Kageyama Y;
 XX Shikata S, Nomura M;
 XX WPI, 1999-287736/27.
 DR P-PSDB; AAY17088, AAY17090.
 XX Alkali protease from *Bacillus* used in washing powders.
 XX Disclosure; Page 58-63; 71pp; Japanese.
 PS The invention relates to alkaline proteases produced by strains of
 CC *Bacillus*. The proteases ability to digest casein is not inhibited by
 CC oleic acid and they have a high stability to oxidising agents. The
 CC alkaline protease of the invention has the following properties: (a) it
 CC is active over the pH range 4-13 and has at least 80% of its optimum
 CC activity over the range pH 6-12; (b) after 30 minutes at 40 deg. C it is
 CC stable over the pH range 6-11; c) its isoelectric point is 8.9-9.1; (d)
 CC its ability to digest casein is not inhibited by oleic acid; (e) it has
 CC molecular weight about 43,000 by SDS-PAGE. The alkaline proteases can be
 CC used as enzymes in washing compositions for use in automatic dishwashers
 CC and for washing clothes. The stability to oxidising agents allows the
 CC enzyme to be an effective component of washing compositions including
 CC bleaches. The present sequence represents an alkaline protease encoding
 CC DNA. (Updated on 20-MAR-2003 to correct DR field.)
 XX
 SQ Sequence 1923 BP; 576 A; 418 C; 475 G; 454 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 3,03e-165 Length: 1923
 Score: 2191.00 Matches: 419
 Percent Similarity: 99.54% Conservative: 13
 Best Local Similarity: 95.54% Mismatches: 2
 Query Match: 97.59% Indels: 0
 DB: 2 Gaps: 0

US-09-985-689A-2 (1-434) x AAX37278 (1-1923)

QY 1 AsnAspValAlaArgGlyIleValIysAlaAspValAlaGlnSerSerThrGlyLeuTyr 20
 Db 619 AATGATGTTGCGGTGGGATTTGTCAGAGGGATGTTGGCTCAGAGCAGCTACGGGTGTAT 678
 QY 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
 Db 679 GGACAAAGGACAGATCGTAGCGTTGCCGATACAGGCGCTTGATACAGGTGCGCAATGACAGT 738
 QY 41 SerMetHisGluAlaPheArgGlyIleThrAlaLeuTyrAlaLeuGlyValArgThrAsn 60
 Db 739 TCGATGCGATGAAGCTTCGCGGGGAAATTAATCTGCATTATATGCTTGGACGGAGAT 798
 QY 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
 Db 799 AATGCCAATGATACGAATGCTCATGTAGCGATGTGGCTGGCTCGGCTATTAGGAAACGCGC 858
 QY 81 AlaThrAsnIleGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
 Db 859 TCCACTAAATAAGAAATGGCGCTCAGGCGAATCTAGTCTTCATCTCATGTATGATAGC 918
 QY 101 SerGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaPheSer 120
 Db 919 GGTGGGGGACTTGGAGGACTACCTCGAATCTGCAAACTTATTCAGCCCAAGCATACAGT 978
 QY 121 AlaGlyValArgIleHisThrAsnSerTrpGlyValAlaValAsnGlyValAlaThrThr 140
 Db 979 GCTGGTGCCAGAAATTCATAAACTCCTGGGGAGCAGCAGTGAATGGGCTTACACAACA 1038
 QY 141 AspSerArgAsnValAspAspTyrValArgIleAsnAspMetThrIleLeuPheAlaAla 160
 Db 1039 GATTCAGAAATGTGGATGACTATGTGCGCAAAATGATATGACGATCCTTTTCGCTGCC 1098

QY 161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaIysAsnAlaIle 180
 Db 1099 GGAATGAAGGACCAACGCGGAAACCATCATGTCAGCCAGGACAGCTAAATAATGCAATA 1158
 QY 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerThrAlaAspAsnIleAsn 200
 Db 1159 ACAGTCGGAGCTACGGGAAACCTCCGCCCAAGCTTTGGGTCTTATCGGACAAATATCAAC 1218
 QY 201 HisValAlaGlnPheSerSerArgGlyProThrIlyAspGlyArgIleIysProAspVal 220
 Db 1219 CATGTGGCAGATTCTTTCACGTGGACGCAAGAGATGGACGATCAACCCGGATGTC 1278
 QY 221 MetAlaProGlyThrThrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
 Db 1279 ATGGCACCGGGAACGTTTATCATGACCAAGATCTTCTTCTGACCGGATTCCTCTTC 1338
 QY 241 TrpAlaAsnHisAspSerIlyValAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260
 Db 1339 TGGCGCAACCATGACAGTAATATGATCATCATGGTGGACGCTCCATGGCTACACCGATC 1398
 QY 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValIysAsnArgIleThrPro 280
 Db 1399 GTTGCTGGAACGTTGGCACAGCTTCGTGAGCATTTTGTGAAAAACAGAGGCATCACACA 1458
 QY 281 LysProSerLeuLeuIlyAlaIleAlaGlyAlaAlaAspValGlyLeuGlyTyr 300
 Db 1459 AAGCCTTCTTATTAAGCGGACATGATGCGGTGCGAGCTGACATCGCCTTGGCTAC 1518
 QY 301 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerIleAsnValAlaTyr 320
 Db 1519 CCGAACGGTAACCAAGGATGGGACGAGTGACATGGATAAATCCCTGAACGTTGGCTAT 1578
 QY 321 ValAsnGluSerSerAlaLeuSerThrSerGlnIlyAlaThrThrThrPheThrAlaThr 340
 Db 1579 GTGACGAGTCCAGTTCTTATCCACCGCCAAAAGCGACGCTACTCGTTTACTGTCTACT 1638
 QY 341 AlaGlyLysProLeuIlyIleSerLeuValTrpSerAspAlaProAlaSerThrThrAla 360
 Db 1639 GCCGGCAAGCCTTTGAAATCTCCCTGGTATGGTCTGATGCCCTTCGAGGACCAACTGCT 1698
 QY 361 SerValThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrArgTyr 380
 Db 1699 TCGGTAAACGTTGTCATGATCTGGACCTTGTCATTACCGCTCCCAATGGCACAGTAT 1758
 QY 381 ValGlyAsnAspPheSerAlaProPheAspAsnAsnTrpAspGlyArgAsnAsnValGlu 400
 Db 1759 GTAGGAATGACTTTACTTCGCCATACATGATACTGGATGGCCGCAATACGTAGAA 1818
 QY 401 AsnValPheIleAsnSerProGlnSerGlyThrThrIleGluValGlnAlaTyrAsn 420
 Db 1819 AATGTATTTTAAATGACCAACAAAGCGGACGATATACAATTGAGGTACAGGCTTATAAC 1878
 QY 421 ValProValGlyProGlnAsnPheSerLeuAlaIleValAsn 434
 Db 1879 GTACCGTTGGACACAGACTTCTGTTGGCAATTGTGAAT 1920

RESULT 4
 AAV82382
 ID AAV82382 standard; DNA; 3003 BP.
 XX
 AC AAV82382;
 XX
 DT 12-APR-1999 (first entry)
 XX
 DE *Bacillus* JPI170 protease gene.
 XX
 KW Protease; detergent; surfactant; leather processing; debittering;
 KW flavour; ss.
 XX
 OS *Bacillus* sp.
 XX
 PH Key Location/Qualifiers
 CDS 846..2771

FT /*tag= a
 FT 846..944
 FT /*tag= b
 FT 1470..2768
 FT /*tag= c
 FT
 FX
 FN

WC9856927-A2.

17-DEC-1998.

09-JUN-1998; 98WO-US012005.

12-JUN-1997; 97US-00873479.

PA (NOVO) NOVO NORDISK BIOTECH INC.

PI Sloma A, Christianson L;

DR WPI; 1999-080908/07.

DR P-PSDB; AAW89547.

XX Novel protease from *Bacillus subtilis* LC20 - useful in laundry and
 PT dishwashing detergents and for leather processing.
 XX

PS Claim 11; Page 52-53; 77pp; English.

XX This nucleotide sequence encodes a novel protease (see AAW82382) of
 CC *Bacillus* sp. JF170 (NCBI 12513). The sequence in plasmid p170BAN is
 CC contained in *Bacillus subtilis* LC20 NRRL B-21680. The protease gene was
 CC isolated from chromosomal DNA of JF170 following preparation of probes
 CC based on protease N-terminal and internal peptides (see AAW89549-50),
 CC screening of chromosomal libraries, isolation of the 3' end of the gene
 CC by inverse PCR (see AAW82410-11), reconstruction of 5' and 3' ends and
 CC PCR amplification (see AAW82412-16). Claimed recombinant host cells can
 CC be used in a method for producing the protease. The protease is used in
 CC laundry and dishwashing detergents, for institutional and industrial
 CC cleaning, and for leather processing, as well as for debittering and
 CC enhancing the degree of hydrolysis of protein hydrolysates, for flavour
 CC development through hydrolysis of proteins, degradation of undesired
 CC peptides and in enzymatic synthesis of peptides. It has enhanced
 CC stability towards oxidation under alkaline conditions, e.g. towards
 CC bleaching agents of the peroxy type. The invention also provides mutant
 CC cells in which the protease activity is diminished. Such cells can be
 CC used for the production of heterologous recombinant proteins
 XX

SQ Sequence 3003 BP; 982 A; 504 C; 645 G; 872 T; 0 U; 0 Other;

Alignment Scores:

Prod. No.: 3.14e-159 Length: 3003
 Score: 2118.50 Matches: 405
 Percent Similarity: 97.47% Conservative: 18
 Best Local Similarity: 93.32% Mismatches: 10
 Query Match: 94.37% Indels: 1
 DB: 2 Gaps: 1

US-09-985-689A-2 (1-434) x AAW82382 (1-3003)

QY 1 AsnAspValAlaArgGlyLeValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
 Db 1470 AATGAGCGTGGCGTGGCATTTGAAAGCAGACGTCGCACAAAATAACTTTGGCTTATAT 1529
 QY 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
 Db 1530 GGCACAGACAGATTAGCAGTTTGTCTACTTGGGCTTGTATACAGGAAGAAATGACAGT 1589
 QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
 Db 1590 TCGATGCATGAACATTCCGCGTAAAGATTACCGCACTATATGCACTGGCGCAGAACGAAT 1649
 QY 61 AsnAlaAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
 Db 1650 AACGCCAATGATCCAAATGGACATGGAAACCCATGTTGCTGGATCTGTGTAGGAAT--- 1706

RESULT 5

QY 81 AlaThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
 Db 1707 GCTACAAATAAAGGATGGCACCGCAAGCAACTAGTCTTTCAATCTATTATGGATAGT 1766
 QY 101 SerGlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaPheSer 120
 Db 1767 GGTGGAGGGCTGGAGGACTACCTGCTAATCTCAAAACATTATTAGTCAAGCATATAGT 1826
 QY 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyValAlaAlaValAsnGlyAlaTyrThr 140
 Db 1827 GCTGGAGCGAGAATTCATACGAATTCATGGGGGCTCCAGTAAACGGTGCCTATACACA 1886
 QY 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
 Db 1887 GACTCTCGAATGTTGATGATTATGAGAAAAAATGATATGACGATTTCTTTTGGGGCC 1946
 QY 161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
 Db 1947 GGAATCGAGGACCAAGTACGGTACATCAGTGCACAGGACACAGAAAAATGCGAAT 2006
 QY 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
 Db 2007 ACAGTTGGGGCAACCGAAAAACCTACGTCCAAGCTTCGGATCTTTATGCGGATAATATAAC 2066
 QY 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
 Db 2067 CATGTTGCTCAATTCCTTCACGAGTCTTACTAGAGATGAGCGTATTAAAGCGGAGTC 2126
 QY 221 MetAlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
 Db 2127 ATGCAACAGGATAGTATATCTCTCTGCTAGATCATCATTAGCTCCAGATTCCTCATTC 2186
 QY 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260
 Db 2187 TGGCAAAACCATGATAGTAAATATGCTACATGGTGGTACTTCTATGGCTACTCCAAAT 2246
 QY 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
 Db 2247 GTAGCAGGTAATGTTGCACAAATTAAGGAGCAATTTGTGAAAAATAGAGGGTAACCTCT 2306
 QY 281 LysProSerLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspValGlyLeuGlyTyr 300
 Db 2307 AAGCTTTCCTCTTTAAAGCTGCTTTAAATGCGAGTGTCTCGGATGTGGACTTGGCTTT 2366
 QY 301 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
 Db 2367 CCAATGTTAAACCCCTTAAAAATATCATTGTTGGTCAGATGCACCGTAGCAGCGCA 2546
 QY 321 ValAsnGluSerSerAlaLeuSerThrSerGlnLysAlaThrTyrThrPheThrAlaThr 340
 Db 2427 GTGAATGAAACGAGCCCTTTATCAACAAGTCAAAAAGCAACATATTGCTTTACGGCTCAA 2486
 QY 341 AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrAla 360
 Db 2487 GCTGGTAAACCCCTTAAAAATATCATTGTTGGTCAGATGCACCGTAGCAGCGCA 2546
 QY 361 SerValThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrArgTyr 380
 Db 2547 TCACATACTTTAGTGAATGATTAGACTTAGTAACTACTCAGCAAAATGGAACATAATAC 2606
 QY 381 ValGlyAsnAspPheSerAlaProPheAspAsnTrpAspGlyArgAsnAsnValGlu 400
 Db 2607 GTCGGAATGACITTAGCAGCACCGTATGATACAAATGGGATGGCAGAAACACGCGAA 2666
 QY 401 AsnValPheIleAsnSerProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
 Db 2667 AATGTGTTTATCAATGCTCTCAAGCGGAACGTATACAGTCGAAGTCAGGCTTACAAT 2726
 QY 421 ValProValGlyProGlnAsnPheSerLeuAlaIleValAsn 434
 Db 2727 GTACAGTAAGTCCGGAACCTTTTCTTTAGCGATTGTACAT 2768

AAQ27516
 ID AAQ27516 standard; DNA; 1299 BP.
 AC
 XX
 XX
 XX
 DT 05-FEB-1993 (first entry)
 DE
 XX Alkali-protease Ya enzyme gene.
 DE
 XX
 KW Alkali resistance; surface active agent resistance; detergency improver;
 KW ss.
 XX
 XX Bacillus sp. Y.
 OS
 XX
 PH Key Location/Qualifiers
 FT CDS 1..1299
 FT /*tag= a
 XX
 PN JP04197182-A.
 PD 16-JUL-1992.
 XX
 XX 28-NOV-1990; 90JP-00327110.
 PF
 XX 28-NOV-1990; 90JP-00327110.
 PR
 XX (LLOY) LION CORP.
 PA
 XX
 XX WPI; 1992-288440/35.
 DR P-PSDB; AAR26274.
 DR
 XX
 XX DNA coding alkali-protease Ya enzyme - has good alkali and surfactant
 PT resistance and improves detergency.
 PT
 CC Claim 3; Page 2; 17pp; Japanese.
 CC
 CC The sequence is that of the alkali-protease Ya enzyme gene which can be
 CC used in the recombinant production of Ya enzyme. Ya enzyme is excellent
 CC in alkali resistance and surface active resistance and improves
 CC detergency
 CC
 XX
 SQ Sequence 1299 BP; 401 A; 246 C; 297 G; 355 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 4,19e-150 Length: 1299
 Score: 1995.50 Matches: 382
 Percent Similarity: 94.24% Conservative: 27
 Best Local Similarity: 88.02% Mismatches: 24
 Query Match: 89.02% Indels: 1
 DB: 2 Gaps: 1

US-09-985-689A-2 (1-434) x AAQ27516 (1-1299)

Qy 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
 Db 1 AATGATGTAGCAGAGGATAGTAAAGCTGATGTCACAAAACAATTACGGATTATAT 60
 Qy 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
 Db 61 GGACAGGTCAACTAGTGTGCTAGCAGCAGCAGCTTAGTACAGCTGTAACGATAGT 120
 Qy 41 SerMetHisGluAlaPheArgGlyIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
 Db 121 TCTATGCATGAAGCATTCGCGGGAATACACAGCTCTTTACGCGTTAGGAAGAACTAAT 180
 Qy 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
 Db 181 AATGGAGTGAATCCGAATGGCGATGGCACATGTAGCAGGTTCTGACTTGGTAAT--- 237
 Qy 81 AlaThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
 Db 238 GCTTTTAAATAAGGAATCGCTCCGCAAGCTAACTAGTCTTCCCAATCTATTATGATGATGC 297

RESULT 6

AAQ27516

ID AAT85667 standard; DNA; 1977 BP.

XX

AC AAT85667;

Qy 101 SerGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaPheSer 120
 Db AGCGAGGATTAGTGCTTTACCATGCACTTAAATCGTTATTTAGTCAAGTTGGAAT 357
 Qy 121 AlaGlyAlaArgIleHisThrAsnSerTyrGlyAlaAlaValAsnGlyAlaTyrThr 140
 Db GCTGGAGCAAGAATTCATACCTCTTGGGGAGCCCAAGTAATGGAGCGTCACTGCT 417
 Qy 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
 Db AACTCGAGACAGTGGATAGTATGTCGAAATATGATATGATGACGCTACTTTTGCAGCT 477
 Qy 161 GlyAsnGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
 Db GCTAATGAAGTCTTAATTCAGAACAAATAGTCTCCAGGTACAGCAAAATGCTATT 537
 Qy 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
 Db ACGTCTGGCGCAACGGAATACTATCGCCCAAGCTTGGTTCGATAGCAGTAAACCAAT 597
 Qy 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
 Db CATATTGCACAATTTTCATCGAGAGGAGCTACGAGGATGGACGAATTAAGCCTGACGTA 657
 Qy 221 MetAlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
 Db ACAGCTCCTGGACAACTTTATTTATCAGCAGCTTCCTTCCTAGCTCCAGACTCTTCGTT 717
 Qy 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260
 Db TGGCGGAATTATAACAGTAAATACGCTATATGCGCGGTACCTCCATGGCGACACTATT 777
 Qy 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
 Db GTTCAGGGAATGTGGCGCAATTACGTGAGCAATTTATAAAATAGAGGTATTACTCCT 837
 Qy 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspValGlyLeuGlyTyr 300
 Db AAGCTCTTTTAAATAAAGCTGCCTTATCGCTGCTACTGATGTTGGTTAGGATAT 897
 Qy 301 ProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
 Db CCTAGTGTGACCAAGGCTGGGCGGTGTTACTCTAGATAAATCGTTAAATGTAGCGTAT 957
 Qy 321 ValAsnGluSerSerAlaLeuSerThrSerGlnLysAlaThrTyrThrPheThrAlaThr 340
 Db GTCAATGAGCAACTGCAATTAGCCACAGGACAAAAGCAAGTATTCGTTCCAAAGCAAA 1017
 Qy 341 AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrAla 360
 Db GCGGTAACCTTTAAAAATCTCGTTAGTAGACAGATGCTCTCTGGAAGTACAACTGCA 1077
 Qy 361 SerValThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrArgTyr 380
 Db TCTTTATACACTAGTTAATGATTTAGTCTAGTTATTTACTGCTCCGAATGGACAAAATAT 1137
 Qy 381 ValGlyAsnAspPheSerAlaProPheAspAsnTrpAspGlyArgAsnAsnValGlu 400
 Db GTAGGAATGATTTTAGTTATCTTATGATAATAACTGGGATGCTGCGCAAAATGTTGAG 1197
 Qy 401 AsnValPheIleAsnSerProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
 Db AACGATTATTATAACCTCCGCAATCTGGAACGTTATATANTAGGTTCAAGCTATAAT 1257
 Qy 421 ValProValGlyProGlnAsnPheSerLeuAlaIleValAsn 434
 Db GTACCATCTGGCCCAACAGGTTTCTCACTAGCTATCGTACAT 1299


```

XX AC AAX05926;
XX DT 06-MAY-1999 (first entry)
XX DE WO9856926 Seq ID 11.
XX KW Hyperthermostable; protease; thermophilic; bacterium; subtilisin;
XX KW additive; drug; washing agent; foodstuff; chemical synthesis; ds.
XX OS Thermococcus celer.
XX PN WO9856926-A1.
XX PD 17-DEC-1998.
XX PF 04-JUN-1998; 98WO-JP002465.
XX PR 10-JUN-1997; 97JP-00151969.
XX PA (TAKI ) TAKARA SHUZO CO LTD.
XX PI Takakura H, Morishita M, Shimojo T, Asada K, Kato I;
XX DR WPI; 1999-080907/07.
XX DR P-PSDB; AAW94840.
XX PT Recombinant hyperthermostable protease from Pyrococcus furiosus - and
XX PT gene encoding it, for large scale production of the protease for
XX PT industrial use.
XX PS Disclosure; Page 53-54; 82pp; Japanese.
XX CC The invention relates to a hyperthermostable protease derived from a
XX CC thermophilic bacterium (especially Pyrococcus furiosus). The protease has
XX CC working temperature 40-110 deg.C (optimum 80-95 deg.C), working pH 5-10
XX CC (optimum 6-8), and retains more than 90% of its activity after 8 hours at
XX CC 95 deg.C. The invention also provides gene sequences encoding a
XX CC polypeptide of formula SIG-Ala-Gly-Asn-PRO, where SIG is a signal
XX CC peptide from subtilisin, and PRO is the above protease. Host cells
XX CC (especially Bacillus strains) transformed with vectors comprising the
XX CC genes are used for the recombinant production of the protease. The
XX CC hyperthermostable protease which can be prepared in quantity suitable for
XX CC industrial use, can be used as an additive for drugs, washing agents and
XX CC foodstuffs and for chemical synthesis
XX SQ Sequence 1977 BP; 453 A; 659 C; 554 G; 311 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 5,09e-27 Length: 1977
Score: 457.50 Matches: 136
Percent Similarity: 44.66% Conservative: 69
Best Local Similarity: 29.63% Mismatches: 153
Query Match: 20.38% Indels: 101
DB: 2 Gaps: 17

US-09-985-689A-2 (1-434) x AAX05926 (1-1977)
Qy 8 VallysalaspvalAlaGlnSerSerTyxGlyLeuTyrgGlyGlnIleValAla 27
Db 433 ATAGGGCCGATACCGCTCGGAACCTCCCTCGGTACGACGCGTGTGGTGTGCC 492
Qy 28 ValAlaaspThrGlyLeuaspThrGlyArgAsnAspSerSerMethisGluAlaPheArg 47
Db 493 ATCTCGATAGGGGTATAGCGCGAAC-----CACCCCGATCTGAAG 534
Qy 48 GlyLysIleThrAlaLeuTyx---AlaLeuGlyArgThrAsnAsnAlaAspThrAsn 66
Db 535 GCGAAGGTCTAGGCTGATCGACGCGGTCAACGGCAGGTCTGACCCCTACGATGCCAG 594
Qy 67 GlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlyValAlaThrAsnLys----- 84
Db 595 GGACACGGAAACCGTTCGGGTATCGTTGCCGGAAACCGCGAGCGGTAACTCCAGTAC 654

```

```

Qy 85 ---GlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerSer 101
Db 655 ATAGCGCTCGCCCGCGCGAAGCTCGTCGGCGTCAAGGTTCTCGGTGCGGACGGTTCG 714
Qy 102 GlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaPheSerAla 121
Db 715 GGAAGCTCTCCACCATCATCGCGGGTGTGACTGGGTCTGCAGACACAGGCAAGTAC 774
Qy 122 GlyAlaArgIle-----HsThrAsnSer 129
Db 775 GGGATAAGGCTCATCAACCTCTCCCTCGGTCTCCAGAGCTCCGACGGAACCGACTCC 834
Qy 130 TrpGlyAlaAlaValAsnGlyAlaValThrThrAspSerArgAsnValAspTyrVal 149
Db 835 CTCAGTCAGCGCTCAACACGCTGGACGCC----- 867
Qy 150 ArgLysAsnAspMetThrIleLeuPheAlaAlaGlyAsnGlyProAsnGlyGlyThr 169
Db 868 -----GGTATAGTAGTCTGCGTCCGCGCAACAGCGGCGCGAACACCTACACC 918
Qy 170 IleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArg 189
Db 919 GTCGCTCACCCGCGCGGAGCAAGTCTATACCTCGGTGCA----- 963
Qy 190 ProSerPheGlySerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGly 209
Db 964 -----GTTGACAGCAACGACAACTCGCAGCTTCTCCAGCAGGGA 1005
Qy 210 ProThrLysAspGlyArgIleLysProAspValMetAlaProGlyThrTyrIleLeuSer 229
Db 1006 CCGACCGCGGAGGAGCTCAACCGGAAGTCTGCGCCCGCGGTGTGACATCATAGCC 1065
Qy 230 AlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsnHisAspSerLysTyrAla 249
Db 1066 CCGCGCGCCAGC-----GGAACGAGCATGGCCACCCCGATAAAGCACTACTACACC 1116
Qy 250 TyrMetGlyGlyThrSerMetAlaThrProIleValAlaGly---AsnValAlaGlnLeu 268
Db 1117 AAGCGCTCTGAAACGAGCATGGCCCGCACGCTTTCGGCGGTGGCGGCTCATCTC 1176
Qy 269 ArgGluHisPheValLysAsnArgGlyIleThrProLys-----ProSerLeuLys 286
Db 1177 CAGGCCAC-----CCGAGCTGGACCCCGGCAAGTGAAG 1212
Qy 287 AlalaLeuIleAlaGlyAla-----AlaAspValGlyLeu 298
Db 1213 ACCGCCCTCATCGAGACCGCGCATAGTCGCCCAAGAGATAGCGGACATCGCTAC 1272
Qy 299 GlyTyrProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeu----- 316
Db 1273 GGTGGC-----GGTAGGTGAACGTCTACAAGGCCATCAAGTAC 1311
Qy 317 ---AsnValAlaTyrValAsnGluSerSerAlaLeuSerThrSerGlnLysAlaThrTyr 335
Db 1312 CACGACTACGCCAAGCTCACCTTCACCGGTCTCCGTCGCCGCAAGGAAAGCGCCACCC 1371
Qy 336 ThrPheThrAlaThrAlaGlyLysProLysIleSerLeuValTrpSerAspAlaPro 355
Db 1372 ACCTTGACGTACGGCGGCCACCTTCGTGACCCGCCCTCTACTCTGGAC----- 1422
Qy 356 AlaSerThrThrAlaSerValThrLeuValAsnAspLeuAspLeuValIleThrAlaPro 375
Db 1423 -----ACGGGTCTCGAGCGACATCGACCTCTACCTCTACGACCCC 1461
Qy 376 AsnGlyThrArgTyrValGlyAsnAspPheSerAlaProPheAsnAsnTrpAspGly 395
Db 1462 AACGGGAACGAG-----GTTGACTCTCTACACGCC 1494
Qy 396 ArgAsnAsnValGluAsnValPheIleAsnSerProGlnSerGlyThrTyrThrIleGlu 415
Db 1495 TACTACGGCTTCGAGAGGTCGGCTACTACAACCCGACCGCGGAACTGAGCGGTCAAG 1554

```

QY 416 ValGlnAlaTyrAsnValProValGlyProGlnAsnPheSerLeuAlaIleValAsn 434
 DB 1555 GTCGACGTACAAG-----GCGCGGGCAACTACCAGGTCCGACGTCGTCGACG 1602

RESULT 8

ID AAX05920 standard; DNA; 1236 BP.
 AC AAX05920;
 DT 06-MAY-1999 (first entry)
 DE Hyperthermostable protease fragment encoding DNA.
 KW Hyperthermostable; protease; thermophilic; bacterium; subtilisin;
 KW additive; drug; washing agent; foodstuff; chemical synthesis; ds.
 OS Pyrococcus furiosus.
 PN W09856926-A1.
 PD 17-DEC-1998.
 PF 04-JUN-1998; 98WO-JF002465.
 PR 10-JUN-1997; 97JP-00151969.
 PA (TAKI) TAKARA SHUZO CO LTD.

PI Takakura H, Morishita M, Shimojo T, Asada K, Kato I;
 XX WPI; 1999-080907/07.
 DR P-PSDB; AAM94836.

XX Recombinant hyperthermostable protease from *Pyrococcus furiosus* - and
 PT gene encoding it, for large scale production of the protease for
 PT industrial use.

XX Claim 6; Page 37-38; 82pp; Japanese.

CC The invention relates to a hyperthermostable protease derived from a
 CC thermophilic bacterium (especially *Pyrococcus furiosus*). The protease has
 CC working temperature 40-110 deg.C (optimum 80-95 deg.C), working pH 5-10
 CC (optimum 6-8), and retains more than 90% of its activity after 8 hours at
 CC 95 deg.C. The invention also provides gene sequences encoding a
 CC polypeptide of formula SIG-Ala-Gly-Gly-Asn-PRO, where SIG is a signal
 CC peptide from subtilisin, and PRO is the above protease. Host cells
 CC (especially *Bacillus* strains) transformed with vectors comprising the
 CC genes are used for the recombinant production of the protease. The
 CC hyperthermostable protease which can be prepared in quantity suitable for
 CC industrial use, can be used as an additive for drugs, washing agents and
 CC foodstuffs and for chemical synthesis

XX SQ Sequence 1236 BP; 368 A; 251 C; 309 G; 308 T; 0 U; 0 Other;

Alignment Scores:

| Pred. No.: | 1,028-24 | Length: | 1236 |
|------------------------|----------|---------------|------|
| Score: | 425.50 | Matches: | 140 |
| Percent Similarity: | 42.77% | Conservative: | 61 |
| Best Local Similarity: | 29.79% | Mismatches: | 147 |
| Query Match: | 18.95% | Indels: | 122 |
| DB: | 2 | Gaps: | 19 |

US-09-985-689A-2 (1-434) x AAX05920 (1-1236)

QY 12 ValAlaGlnSerSerTyrGly-LeuTyr-----GlyGlnGlyG1 24
 DB 24 GTCTCAGCTCAAGTATATGCAACTTACGTTTGGAACTTGGGATATGATGTTCTCGAAT 83
 QY 24 nileValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisG1 44
 DB 84 CACATAGATATATGACACTGGATTGAC-----GCTTCTCATCC 125

QY 44 uAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAsn-- 63
 DB 126 AGATCTCCCAAGGAAAGTA-----ATTGGTGGTAGATTTCGTCATG 170
 QY 64 -----AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuG1 78
 DB 171 TAGGAGTTATCCATAGCAGCATGCACTCATGTAGCTTCAATAGCAGCTGG 230
 QY 78 YAsnGlyAlaThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPh 94
 DB 231 TACTGGAGCAGCAAGTAATGGCAAGTACAAAGGAATGGCTCCAGGAGCTAAGCTGGCGG 290
 QY 94 eGlnSerIleMet-----AspSerSerGlyGlyLeuGlyGlyLeuProSerAsnLeuG1 112
 DB 291 AATAAGGTTCTAGTGGCCGATGGTTCTCGAAGCATATCTACTATAAATAAGGGAGTTGA 350
 QY 112 nThrLeuPheSerGlnAlaPheSerAlaGlyAlaArgIleHisThrAsnSerTyrGlyAl 132
 DB 351 GTGGCGCGTTGATAACAAAGATAAGTACGGAATTAAGTTCATTAATCTTCTTGGTTC 410
 QY 132 a-----AlaValAsnGlyAlaTyrThrTh 140
 DB 411 AAGCCAGAGCTCAGATGCTAGCTCTAAAGTCAGGCTGTTAATGCAGCGTGGGATGC 470
 QY 140 rAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAl 160
 DB 471 T-----GGATTAGTTGTTGTTGTTGCCG 494
 QY 160 aGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaI 180
 DB 495 TGGAAACAGTGGACCTAACAAATGATACAAATCGTTCTCCAGCAGCTGCAGCAAGTTAT 554
 QY 180 eThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAs 200
 DB 555 TACAGTTGGAGCC-----GTTGCAAGATATGA 581
 QY 200 nHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVa 220
 DB 582 TGTATATAACAAGCTTCTCAAGCAGAGGGCCAACTGCAGACGGCAGCTTAAGCTTGAGT 641
 QY 220 lMetAlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPh 240
 DB 642 TGTGTCTCCAGAAACTGGATAATTGCTGCCAGACCAAGT-----GGAACATGACAT 692
 QY 240 eTTPAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProI 260
 DB 693 GGCTCAACCAATTAATGACTATTACACACAGCTCTCTGGACATCATGCGCACTCTCA 752
 QY 260 eValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPr 280
 DB 753 CGTAGCTGTATTGACAGCCCTCTTGTCTCAA-----GCACACCC 791
 QY 280 cLys-----ProSerLeuLysAlaAlaLeuIleAlaGlyAla----- 293
 DB 792 GAGCTGGACTCCAGACAAAGTAAACAGCCCTCATAGNAACCTGTGATATCGTAAGCC 851
 QY 294 -----AlaAspValGlyLeuGlyTyrProAsnGlyAsnGlnGlyTyrGlyArgVa 310
 DB 852 AGATGAATATAGCGGATATAGCCTACCGTGCA-----GGTAGGCT 890
 QY 310 lThrLeuAspLysSerLeuAsnValAlaTyrValAsnGluSerSerAlaLeuSerSerSe 330
 DB 891 TAATGATACACAGGCTATTAAC-----TACGATAACTATGCAAGCTAGTGTTCACCTGG 944
 QY 330 rGlnLysAla-----ThrTyrThrPheThrAlaThrAlaGlyLysProLe 345
 DB 945 ATATGTGTGCAACAAAGGAGCAGCAAACTCACCAGTTCGTTATTAGCGAGCTTCGTCGT 1004
 QY 345 uLysIleSerLeuValTyrSerAspAlaProAlaSerThrThrAlaSerValThrLeuVa 365
 DB 1005 AACTGCCCATATTACTGGACAAATGCCAAT----- 1035
 QY 365 lAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrArgTyrValGlyAsnAspPh 385

```

Db 1036 AGCGACCTTGATCTTTACTCTACGATCCCAATGGAACACG----- 1077
QY 385 sSerAlaProPheAspAsnAsnTyrAspGlyArgAsnAsnValGluAsnValPheIleAs 405
Db 1078 -----GTTGACTACTCTTACACCGCTTACTATGATTCGAAAGGTTGGTTATTA 1127
QY 405 nSerProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsnValProValGlyPr 425
Db 1128 CAAACCACTGATGGAACATGACAAATTAAGTTGTAACTACAGC-----GGAG 1178
QY 425 oGlnAsnPheSerLeuAlaIleValAsn 434
Db 1179 TGCAAACTATCAAGTAGATGTGTAAGT 1206

RESULT 9
AAT85668
ID AAT85668 standard; DNA; 1566 BP.
XX
AC AAT85668;
XX
DT 17-OCT-2003 (revised)
DT 20-APR-1998 (first entry)
XX
DE Pyrococcus furiosus protease coding sequence.
XX
KW Protease; research reagent; thermal stability; pyrococcus furiosus; ss.
XX
OS Pyrococcus furiosus; DSM-3638.
XX
FH Key Location/Qualifiers
FT CDS 1..1566
FT /tag= a
FT /transl_except= (pos: 1282..1284, aa: Xaa)
FT /note= "Xaa= Gly, Val"
XX
PN WO9721823-A1.
XX
PD 19-JUN-1997.
XX
PF 07-NOV-1996; 96WO-JP003253.
XX
PR 12-DEC-1995; 95JP-00323285.
XX
PA (TAKI ) TAKARA SHUZO CO LTD.
XX
PI Takakura H, Morishita M, Yamamoto K, Mitta M, Asada K;
PI Tsunasawa S, Kato I;
XX
DR WPI; 1997-332794/30.
DR P-PSDB; AAW24122.
XX
PT Protease(s) and genes encoding them obtained from Thermococcus and
PT Pyrococcus strains - have extremely high thermal stability and are useful
PT industrially and as research reagents.
XX
PS Claim 7; Page 90-91; 159pp; Japanese.
XX
CC This sequence represents the coding sequence for the protease from
CC Pyrococcus furiosus DSM-3638. This sequence encodes a protease of the
CC invention. The processes of the invention have extremely high thermal
CC stability. The proteases can be used as research reagents, and
CC industrially in the food, drug and chemical industries. (Updated on 17-
CC OCT-2003 to standardise OS field)
XX
SQ Sequence 1566 BP; 467 A; 342 C; 372 G; 384 T; 0 U; 1 Other;

Alignment Scores:
Pred. No.: 1,36e-24 Length: 1566
Score: 425.50 Matches: 140
Percent Similarity: 42.77% Conservative: 61
Best Local Similarity: 29.79% Mismatches: 147
Query Match: 18.95% Indels: 122

```

```

DB: 2 Gaps: 19
US-09-985-689A-2 (1-434) x AAT85668 (1-1566)
QY 12 ValAlaGlnSerSerTyrGly-LeuTyr-----GlyGlnGlyCl 24
Db 24 GTCTGAGCTCAAGTTATGCGCACTTACCTTTGGAACCTGGGATATGATGCTTCTGGAAT 83
QY 24 nIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisG1 44
Db 84 CACAATAGGAATTAATGACACTGGAAATGAC-----GCTTCTCATCC 125
QY 44 uAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAlaAsn-- 63
Db 126 AGATCTCCAAAGGAAAAGTA-----ATTGGTGGGTAGATTTTGTCAATGG 170
QY 64 -----AspThrAsnGlyHisGlyThrHisValAlaGlySerValIleuG1 78
Db 171 TAGGAGTTATCCATACGATGACCATGGACATGGAATCATGACTCATGAGTTCATAGCAGCTGG 230
QY 78 YAsnGlyAlaThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPh 94
Db 231 TACTGGAGCAGCAAGTAATGGCAAGTACAAAGGAATGGCTCCAGGAGCTAAGCTGGCGG 290
QY 94 eGlnSerIleMet-----AspSerSerGlyGlyLeuGlyGlyLeuProSerAsnLeuG1 112
Db 291 AATTAAGTTCTAGGTGCGCATGTTCTGGAAGCATATCTACTATTAATTAAGGAGTTGA 350
QY 112 nThrLeuPheSerGlnAlaPheSerAlaGlyAlaArgIleHisThrAsnSerTyrGlyAl 132
Db 351 GTGGCGCTTGATACAAAGATAAAGTACGAATTAAGTCAATTAATCTTCTTCTGTTTC 410
QY 132 a-----AlaValAsnGlyAlaTyrThr 140
Db 411 AAGCAGAGCTCAGATGGTACTGACCTCTAAGTCAAGCTGTTTAATGCGGCGGATGC 470
QY 140 rAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAla1 160
Db 471 T-----GGATTAGTTGTTGTTGCTGCCGC 494
QY 160 aGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAla1 180
Db 495 TGGAAACAGTGGACCTTAACAAAGTATACAATCGGTTCTCCAGCAGCTGCAAGCAAGTTAT 554
QY 180 eThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAs 200
Db 555 TACAGTTGGAGCC-----GTTGACAAAGTATGA 591
QY 200 nHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVa 220
Db 582 TGTATTAACAAGCTTCTCAAGCAGAGGGCCAACTGCAGACGGCAGGCTTAAGCTGAGGT 641
QY 220 lMetAlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPh 240
Db 642 TGTGCTCCAGGAACCTGGATTAATGCTGCCAGAGCAAGT-----GGAAGTACAT 692
QY 240 eTTPAlaAsnHisAspSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProI1 260
Db 693 GGGTCAACCAATTAATGACTATTACACAGCAGCTCTCTGGGACATCAATGGCAATCTCTCA 752
QY 260 eValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPr 280
Db 753 CGTAGCTGGTATTCCAGCCCTCTTGCTCCAA-----GCACACCC 791
QY 280 oLys-----ProSerLeuLeuLysAlaAlaLeuIleAlaGlyAla----- 293
Db 792 GAGCTGGACTCCAGACAAAGTAAAAACAGCCCTCATAGAACTGCTGATATCGTAAAGCC 851
QY 294 -----AlaAspValGlyLeuGlyTyrProAsnGlyAsnGlnGlyTyrGlyArgVa 310
Db 852 AGATGAATACCCATATAGCTTACGCTGCA-----GGTAGGT 890
QY 310 lThrLeuAspLysSerLeuAsnValAlaTyrValAsnGluSerSerAlaLeuSerThrSe 330

```

Db 891 TAATGATACAGGCTATAAC-----TACGATAACTATGCAAGAGTAGTGTTCACCTGG 944
 Qy 330 rglNlysAla-----ThrTyrThrPheThrAlaThrAlaGlyLysProLe 345
 Db 945 ATATGTTGCCACAAAGGAGGAGCAAACTCACAGATTCGTTATTAGCGAGGTTCGTTGCT 1004
 Qy 345 ulysileSerLeuValTyrSerAspAlaProAlaSerThrThrAlaSerValThrLeuVa 365
 Db 1005 AACTGCCACATTATCTGGGCAATGCCAAT----- 1035
 Qy 365 lIasnAspLeuAspLeuValIleThrAlaProAsnGlyThrArgTyrValGlyAsnAspPh 385
 Db 1036 -AGCGACCTTGATCTTTACTCTAGCATCCCAATGGAACACAG----- 1077
 Qy 385 eserAlaProPheAspAsnAspTyrAspGlyArgAsnAsnValGluAsnValPheIleAs 405
 Db 1078 -----GTGACTACTCTTACCCCTACTATGATTCGAAAGGTGGTTATTA 1127
 Qy 405 sSerProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsnValProValGlyPr 425
 Db 1128 CAACCCAACTGATGGAACATGACAAATTAAGGTTGTAAGCTACAGC-----GGAAG 1178
 Qy 425 cglNAsnPheSerLeuAlaIleValAsn 434
 Db 1179 TCGAAACTATCAAGTAGATGTGTAAGT 1206

RESULT 10

AAT85695

ID AAT85695 standard; DNA; 1962 BP.

AC AAT85695;

XX 17-OCT-2003 (revised)

DT 20-APR-1998 (first entry)

XX Pyrococcus furiosus PFUS protease coding sequence.

XX Protease; research reagent; thermal stability; pyrococcus furiosus; ss.

XX Pyrococcus furiosus; DSM-3638.

XX WO9721823-A1.

XX 19-JUN-1997.

XX 07-NOV-1996; 96WO-JP003253.

XX 12-DEC-1995; 95JP-00323285.

XX (TAXI) TAKARA SHUZO CO LTD.

XX Takakura H, Morishita M, Yamamoto K, Mitta M, Asada K;

XX Tsunashawa S, Kato I;

XX WPI; 1997-332794/30.

XX P-PSDB; AAW24129.

XX Protease(s) and genes encoding them obtained from Thermococcus and

XX Pyrococcus strains - have extremely high thermal stability and are useful

XX industrially and as research reagents.

XX Disclosure; Page 123-125; 159pp; Japanese.

XX This sequence represents the coding sequence for the protease from

XX Pyrococcus furiosus DSM-3638. This sequence encodes a protease of the

XX invention. The proteases of the invention have extremely high thermal

XX stability. The proteases can be used as research reagents, and

XX industrially in the food, drug and chemical industries. (Updated on 17-

XX OCT-2003 to standardise OS field)

XX Sequence 1962 BP; 602 A; 399 C; 471 G; 490 T; 0 U; 0 Other;

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Alignment Scores:
 Pred. No.: 1.8e-24 Length: 1962
 Score: 425.50 Matches: 140
 Percent Similarity: 42.77% Conservative: 61
 Best Local Similarity: 29.79% Mismatches: 147
 Query Match: 18.95% Indels: 122
 DB: 2 Gaps: 19

US-09-985-689A-2 (1-434) x AAT85695 (1-1962)

Qy 12 ValAlaGlnSerSerTyrGly-LeuTyr-----GlyGlnGlyG1 24
 Db 420 GTCTGCAGCTCAAGTTATGTCACAACTTACGTTTGGAACTTGGGATATGATGGTTCTGGAAT 479
 Qy 24 nIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisG1 44
 Db 480 CACAATAGGAATAATGTCACACTGGAATGAC-----GCTTCTCATCC 521
 Qy 44 uAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAsn-- 63
 Db 522 AGATCTCCAAGGAAAAGTA-----ATTGGTGGGTAGATTGTTGTCATGCG 566
 Qy 64 -----AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuG1 78
 Db 567 TAGAGTTATCCATACGATGACCATGACATGGAATCATGTAGTTCATATAGCAGCTGG 626
 Qy 78 YAsnGlyAlaThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPh 94
 Db 627 TACTGGGCGCAGCAAGTAATGGCAAGTACAGGAATGGCTCCAGGAGCTAAGCTGGCGGG 686
 Qy 94 eGlnSerIleMet-----AspSerSerGlyGlyLeuGlyGlyLeuProSerAsnLeuG1 112
 Db 687 AATTAAGGTTCTAGGTGCGCATGTTCTCGAAGCATATCTACTATAAATTAAGGGATTGA 746
 Qy 112 nThrLeuPheSerGlnAlaPheSerAlaGlyAlaArgIleHisThrAsnSerTyrGlyAl 132
 Db 747 GTGGGCGGTTGATAACAAAGATAAGTACGAATTAAGTTCATTAATCTTCTTCTGTTTC 806
 Qy 132 a-----AlaValAsnGlyAlaTyrThrTh 140
 Db 807 AAGCCAGAGCTCAGATGCTAGCTCTAAGTCTAGGCTGTGTAATGTCAGCTGGGATGC 866
 Qy 140 rAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAl 160
 Db 867 T-----GGATTAGTTGTTGTTGTTGCTTCC 890
 Qy 160 aGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaI 180
 Db 891 TGGAAACAGTGGACCTTACCAAGTATACAAATCGTTCTCCAGCAGCTGCAAGCAAAAGTTAT 950
 Qy 180 eThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAs 200
 Db 951 TACAGTTGGAGCC-----GTTGACCAAGTATGA 977
 Qy 200 nHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVa 220
 Db 978 TGTATAAACAAGCTTCTCAAGCAGAGGGCCAACTGCAGACGCGCAGGCTTAAGCCTGAGGT 1037
 Qy 220 lMetAlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPh 240
 Db 1038 TGTGTCTCCAGAAATGGAATAATGTCTCCAGACCAAGT-----GGAACCTAGCAT 1088
 Qy 240 eTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProI 260
 Db 1089 GGGTCAACCAATTAATGACTATTACACAGCAGCTCTGGGACATCAATGGCACTCTCTCA 1148
 Qy 260 eValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPr 280
 Db 1149 CGTAGCTGGTATTGACGCCCTCTGTCTCCAA-----GCACACCC 1187
 Qy 280 oLys-----ProSerLeuLeuLysAlaAlaLeuIleAlaGlyAla----- 293
 Db 1188 GAGCTGCACTCCAGACAAAGTAAACAGCCCTCATAGAACTGCTGATATCGTAAAGCC 1247

QY 294 -----AlaAspValGlyLeuGlyTyrProAsnGlyAsnGlnGlyTrpGlyArgVa 310
 Db 1248 AGATGAATAGCCGATATAGCCACGGTGCA-----GGTAGGGT 1286
 QY 310 lThrLeuAspLysSerLeuAsnValAlaTyrValAsnGlnSerSerAlaLeuSerThrSe 330
 Db 1287 TAATGCATACAGCGCTATAAC-----TACGATAACTATGCAAGCTAGTGTCTACTGG 1340
 QY 330 rGlnLysAla-----ThrTyrThrPheThrAlaThrAlaGlyLysProLe 345
 Db 1341 ATATGTTGCCAACAAAGCGACCAACTCACCAGTTTCGTTATTAGCGGAGCTTCGTTGGT 1400
 QY 345 uLysLeSerLeuValTrpSerAspAlaProAlaSerThrThrAlaSerValThrLeuVa 365
 Db 1401 AACTGCCACATTATCTAGCGACATGCCAAT-----TACGATAACTATGCAAGCTAGTGTCTACTGG 1431
 QY 365 lAsnAspLeuAspLeuValLleThrAlaProAsnGlyThrArgTyrValGlyAsnAspPh 385
 Db 1432 -AGGACCTTGAATCTTACTCTACGATCCCAATGGAAACAG-----TACGATAACTATGCAAGCTAGTGTCTACTGG 1473
 QY 385 eSerAlaProPheAspAsnAsnTrpAspGlyArgAsnAsnValGluAsnValPheLeAs 405
 Db 1474 -----GTTGACTACTCTTACACCGGCTACTATGGATTCCGAAAGGTTGGTTATTA 1523
 QY 405 nSerProGlnSerGlyThrTyrThrTrileGluValGlnAlaTyrAsnValProValGlyPr 425
 Db 1524 CAACCCAACTGATGGAACATGGCAATTAAGTTTGAAGTACAGC-----GGAAG 1574
 QY 425 cGlnAsnPheSerLeuAlaLleValAsn 434
 Db 1575 TGCAAACTATCAAGTAGATGGTGAAGT 1602

RESULT 11

AAX05929
 ID AAX05929 standard; DNA; 1962 BP.

AC AAX05929;

DT 06-MAY-1999 (first entry)

DE Hyperthermostable protease encoding DNA.

KW Hyperthermostable; protease; thermophilic; bacterium; subtilisin; additive; drug; washing agent; foodstuff; chemical synthesis; ds.

OS Pyrococcus furiosus.

PN WO9856926-A1.

PD 17-DEC-1998.

PF 04-JUN-1998; 98WO-JP002465.

PR 10-JUN-1997; 97JP-00151969.

XX (TAKI) TAKARA SHUZO CO LTD.

PI Takakura H, Morishita M, Shimojo T, Asada K, Kato I;

DR WPI; 1999-080907/07.

DR P-PSDB; AAW94841.

XX Recombinant hyperthermostable protease from Pyrococcus furiosus - and gene encoding it, for large scale production of the protease for industrial use.

PS Disclosure; Page 59-60; 82pp; Japanese.

CC The invention relates to a hyperthermostable protease derived from a thermophilic bacterium (especially Pyrococcus furiosus). The protease has working temperature 40-110 deg.C (optimum 80-95 deg.C), working pH 5-10 (optimum 6-8), and retains more than 90% of its activity after 8 hours at

CC 95 deg.C. The invention also provides gene sequences encoding a polypeptide of formula SIG-Ala-Gly-Asn-PRO, where SIG is a signal peptide from subtilisin, and PRO is the above protease. Host cells (especially Bacillus strains) transformed with vectors comprising the genes are used for the recombinant production of the protease. The hyperthermostable protease which can be prepared in quantity suitable for industrial use, can be used as an additive for drugs, washing agents and foodstuffs and for chemical synthesis

XX Sequence 1962 BP; 602 A; 399 C; 471 G; 490 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 1.8e-24 Length: 1962
 Score: 425.50 Matches: 140
 Percent Similarity: 42.77% Conservative: 61
 Best Local Similarity: 29.79% Mismatches: 147
 Query Match: 122
 Gaps: 19

US-09-985-689A-2 (1-434) x AAX05929 (1-1962)

QY 12 ValAlaGlnSerSerTyrGly-LeuTyr-----GlyGlnGlyVl 24
 Db 420 GTCTGCAGCTCAAGTTATGGCACTTACGTTGGAACTGGGATATGATGTTCTGAAT 479
 QY 24 nIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGl 44
 Db 480 CACAATAGGAATAATTGACACTGCAATTGAC-----GCTTCTCATCC 521
 QY 44 uAlaPheArgGlyLysLleThrAlaLeuTyrAlaLeuGlyArgThrAsnAlaAsn-- 63
 Db 522 AGATCTCCAAAGAAAGTA-----ATTGGTGGGTAGATTGTCATGCG 566
 QY 64 -----AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGl 78
 Db 567 TAGGAGTTATCCATACGATGACCATGGACATGGAACTCATGTAGCTTCAATAGCAGCTGG 626
 QY 78 yAsnGlyAlaThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPh 94
 Db 627 TACTGCAGCAGCAAGTAATGCAAGTACAGGGAATGGCTCCAGAGTAAGTGGCGGG 686
 QY 94 eGlnSerIleMet-----AspSerSerGlyLeuGlyLeuGlyLeuProSerAsnLeuGl 112
 Db 687 AATTAAGGTTCTAGTGCGCGATGGTTCTGGAAAGCATATCTACTATATAATTAAGGAGATGA 746
 QY 112 nThrLeuPheSerGlnAlaPheSerAlaGlyAlaArgLleHisThrAsnSerTrpGlyAl 132
 Db 747 GTGGGCCGTTTGATAACAAAGATAAGTACCGGAATTAAGTCAATCTTCTCTTGGTTC 806
 QY 132 a-----AlaValAsnGlyAlaTyrThrTh 140
 Db 807 AAGCCAGAGCTCAGATGGTACTGACGCTTAAGTCAGGCTGTTAATGCGGTGGGATGC 866
 QY 140 rAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAl 160
 Db 867 T-----GGATTAGTTGTTGGTGGCGC 890
 QY 160 aGlyAsnGlyProAsnGlyThrLleSerAlaProGlyThrAlaLysAsnAlaIl 180
 Db 891 TGGAAACAGTGGACCTTACCAAGTATCAATCGTTCTCCAGCAGCTGCAAGCAAGATTAT 950
 QY 180 eThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAs 200
 Db 951 TACAGTTGGAGCC-----GTTGACAAGTATGA 977
 QY 200 nHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgLleLysProAspVa 220
 Db 978 TGTATAACAAGCTTCTCAAGCAGAGGGCCAACTCGACGACGAGCTTAAGCTTACCTGAGGT 1037
 QY 220 lMetAlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPh 240
 Db 1038 TGTTGTCTCCAGGAAACTGGATAATTGCTGCCAGAGCAAGT-----GGAACATAGCAT 1088

QY 240 eTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIl 260
Db 1089 GGGTCAACCAATATATGACTATACACAGCAGCTCTGGACATCAATGGCACTCTCA 1148
QY 260 eValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyLeuThrPr 280
Db 1149 CGTAGCTGTATTGCGCCCTCTTCTCCAA-----GCACACCC 1187
QY 280 olys-----ProSerLeuLeuLysAlaAlaLeuileAlaGlyAla----- 293
Db 1188 GAGCTGGACTCCAGCAAAAGTAAACAGCCCTCATAGAACTGCTATATCGTAAGCC 1247
QY 294 -----AlaAspValGlyLeuGlyTyrProAsnGlyAsnGlnGlyTyrpGlyArgVa 310
Db 1248 AGATGAAATAGCGGATATAGCTAGCTGCA-----GGTAGGGT 1286
QY 310 lThrLeuAspLysSerLysValAlaTyrValAsnGlnSerSerAlaLeuSerThrSe 330
Db 1287 TAATGCATACAAAGGTATTAAC-----TACGATAACTATGCAAGCTAGTGTCTACTGG 1340
QY 330 rGlnLysAla-----ThrTyrThrPheThrAlaThrAlaGlyLysProLe 345
Db 1341 ATATGTTGCCAAACAAAGGAGCAAACTACCACTTCGTTATTAGCGAGCTTCGTCGT 1400
QY 345 uLysIleSerLeuValTyrSerAspAlaProAlaSerThrThrAlaSerValThrLeuVa 365
Db 1401 AACTGCCCATATATCTGGGCAATGCCAAT----- 1431
QY 365 lAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrArgTyrValGlyAsnAspPh 385
Db 1432 -AGCAGCCTTGATCTTACCTCTACGATCCCAATGGAACCAAG----- 1473
QY 385 eSerAlaProPheAspAsnAsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAs 405
Db 1474 -----GTTGACTACTCTTACACCCGCTACTATGATTCGAAAGGTGGTATTA 1523
QY 405 nSerProGlnSerGlyThrTyrIleGluValGlnAlaTyrAsnValProValGlyPr 425
Db 1524 CAACCCAACTGATGGAACATGGCAATTAAGTTGTAAGCTACAGC-----GGAG 1574
QY 425 oGlnAsnPheSerLeuAlaIleValAsn 434
Db 1575 TGCAAACTATCAAGTAGATGTGTGAAGT 1602

RESULT 12

AA85669

ID AAT85669 standard; DNA; 1977 BP.

XX AAT85669;

AC AAT85669;

DT 20-APR-1998 (first entry)

XX Protease coding sequence.

XX Protease; research reagent; thermal stability; ss.

XX Synthetic.

XX WO9721823-A1.

XX 19-JUN-1997.

XX 07-NOV-1996; 96WO-JP003253.

XX 12-DEC-1995; 95JP-00323285.

XX (TAKI) TAKARA SHUZO CO LTD.

XX Takakura H, Morishita M, Yamamoto K, Mitta M, Asada K;

XX Tsunashawa S, Kato I;

XX WPI; 1997-332794/30.

XX P-PSDB; AAW24123.

XX

PT Protease(s) and genes encoding them obtained from Thermococcus and
PT Pyrococcus strains - have extremely high thermal stability and are useful
PT industrially and as research reagents.

XX Claim 11; Page 95-97; 159pp; Japanese.

XX This sequence represents the coding sequence for a protease of the
CC invention. The proteases of the invention have extremely high thermal
CC stability. The proteases can be used as research reagents, and
CC industrially in the food, drug and chemical industries

XX SQ Sequence 1977 BP; 527 A; 562 C; 513 G; 375 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 1,37e-23 Length: 1977
Score: 414.50 Matches: 134
Percent Similarity: 43.07% Conservative: 68
Best Local Similarity: 28.57% Mismatches: 147
Query Match: 18.46% Indels: 120
DB: 2 Gaps: 19

US-09-985-689A-2 (1-434) x AAT85669 (1-1977)

QY 12 ValAlaGlnSerSerTyrGly-LeuTyr-----GlyGlnGlyGl 24
Db 420 GTCTGCAGCTCAAGTATTATGGCAACTTACGTTTGGAACTTGGGATATGATGGTTCTGGAA 479

QY 24 nIleValAlaValAlaAspThrGlyLeuAspThrGlyA-GAsnAspSerSerMethHisGl 44
Db 480 CACAATAGTAATATTGACACTGGAATTGAC-----GCTTCTCATCC 521

QY 44 uAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAsn-- 63
Db 522 AGATCTCCAAGGAAAGTA-----ATTGGTGGGTAGATTGTCAATGG 566

QY 64 -----AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGl 78
Db 567 TAGAGTTATCCATACGATGACCATGGAATGGAATCACTCATGTAGCTTCAATAGCAGCTGG 626

QY 78 YAsnGlyAlaThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPh 94
Db 627 TACTGGAGCAGCAAGTAATGGCAAGTACAAAGGGAATGGCTCCAGGAGCTAAGCTGCGGG 686

QY 94 eGlnSerIleMet-----AspSerSerGlyLeuGlyGlyLeuProSerAsnLeuGl 112
Db 687 AATTAAGTTCTAGTGCCGATGCTTCTGGAAGCATATCTACTATATAAGGAGTTGA 746

QY 112 nThrLeuPheSerGlnAlaPheSerAlaGlyAlaArgile----- 125
Db 747 GTGGCGGTTGATAACAAGATAAGTACGGAATTAAGTCAATTAATCTTCTTCTGTTTC 806

QY 126 -----HisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyrThrTh 140
Db 807 AGCCAGAGCTCCGACGGAAACCGACTCCCTCAGTCAGCGCTCAACACGCTTGGAGCGC 866

QY 140 rAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAl 160
Db 867 C-----GGTATAGTAGTCTGCTGCGCGC 890

QY 160 aGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIl 180
Db 891 CGGCACAGCGGGCCGCAACACCTACACCGCTCGGCTCACCCTCCGCGCGGAGCAAGTTCAT 950

QY 180 eThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAs 200
Db 951 AACGTCGTTGCA-----GTTGACACCAACGA 977

QY 200 nHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVa 220
Db 978 CAACATGCCAGCTTCTCCAGAGGAGGACCGACCGGAGCGGAAGCTCAAGCGGAAGT 1037

QY 220 lMetAlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerPh 240

QY 215 ArgIleLysProAspValMetAlaProGlyThrTyrlleLeuSerAlaArgSerSerLeu 234
 Db 1045 AGATAAACCCTGACATTCGGCCCGCCAGGATATATATACATGCGCGAAG----- 1095
 QY 235 AlaProAspSerPheTrpAlaAsnHisAspSerLysTyrlleLeuSerAlaArgSerSerLeu 254
 Db 1096 -----GCGAATTCGTAAATGGATATGTAAACATACAGCGGTACA 1134
 QY 255 SerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLys 274
 Db 1135 ACATGGCAACACCTTTGTAGACGAAGTGTCTCTTATGCTTAAC-----GCT 1185
 QY 275 AsnArgGlyIleThrProLysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAla 294
 Db 1186 AATCCAAATCTCACTCCAAATGATGCA-----AAAAATATATATATGTCTACTCAAAA 1239
 QY 295 AspValGlyLeuGlyTyrlleProAsnGlyAsnGlnGlyTyrlleValArgValThrLeuAspLys 314
 Db 1240 AGCTGGGGCCCTCCAGCAAAACGTTGACTATGTTGCGAGGAATTTGATGGCTATGAA 1299
 QY 315 SerLeuAsnValAla----- 320
 Db 1300 GCTATAAGGGTAGCAGGTAATTTAGAGCAATAATATTTGATGTACCAATCATTTATTAT 1359
 QY 321 ValAsnGluSerSerAlaLeuSerThrSerGlnLysAla---ThrTyrllePheThrAla 339
 Db 1360 ATT-----TCAGGTATCTGCGCGGTTCTCGTACAGTATGACTTGGACATTTTAACGCA 1413
 QY 340 Thr---AlaGlyLysProLeuLysIleSerLeuValTrpSerAsp 353
 Db 1414 ACAAAATACAAGTTATCCAAATCGCAATAACATTAATTAATTCCTGAC 1458

RESULT 14

AAAT61454
 ID AAAT61454 standard; DNA; 2539 BP.

AC AAAT61454;

DT 06-OCT-1997 (first entry)

DE Streptomyces viridosporus dhpa gene.

XX asymmetric hydrolase; dhpa; 4-substituted-1,4-dihydropyridine;
 KW derivative; Streptomyces viridosporus; ester; chiral; synthesis;
 KW cardiovascular; treatment; hypertension; ischaemic heart disease; ds.

OS Streptomyces viridosporus.

XX Key Location/Qualifiers
 FH CDS 338..2539

FT /*tag= a

FT /note= "no stop codon given"

FT misc_feature 950..2509

FT /*tag= b

FT /note= "encodes AAW13666"

XX WO9705243-A1.

PN 13-FEB-1997.

XX 30-JUL-1996; 96WO-JP002147.

XX 31-JUL-1995; 95JP-00212975.

PR 29-FEB-1996; 96JP-00067478.

XX (SAOC) MERCIAN CORP.

XX Arisawa A, Matsufuji M, Tsuruta T, Dobashi K, Nakashima T;

PI Isshiki K, Yoshiooka T;

XX WPI; 1997-145682/13.

DR P-PSDB; AAW13666, AAW13667.

XX Asymmetric hydrolase gene derived from Streptomyces viridosporus - acts
 PT on 4-substituted-1,4-di-hydro-pyridine derivatives to produce chiral
 PT derivatives useful for synthesis of cardiovascular drugs.

XX Claim 3; Page 49-55; 78pp; Japanese.

XX This sequence is the Streptomyces viridosporus dhpa gene which encodes an
 CC asymmetric hydrolase which acts on 4-substituted-1,4-dihydropyridine
 CC derivatives. The enzyme allows the efficient conversion of 4-substituted-
 CC 1,4-dihydropyridine esters to chiral partially hydrolysed derivatives,
 CC for use in the synthesis of cardiovascular drugs suitable for the
 CC treatment of e.g. hypertension and ischaemic heart disease

XX Sequence 2539 BP; 433 A; 959 C; 867 G; 280 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 1-12e-18 Length: 2539
 Score: 354.50 Matches: 139
 Percent Similarity: 41.49% Conservative: 56
 Best Local Similarity: 29.57% Mismatches: 173
 Query Match: 15.79% Indels: 103
 DB: 2 Gaps: 17

US-09-985-689A-2 (1-434) x AAT61454 (1-2539)

QY 2 AspValAlaArgGlyIleValLysAlaAspValAlaGlnSerSerTyrlleLeuTyrlleGly 21
 Db 953 GACACGTCCTCGCGGAGATCGCGCCGCCCAAGCGGTGTCGCC---GGCTACGACGGC 1009
 QY 22 GlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSer 41
 Db 1010 AAGGCGTGAAGATCGCGCTCTGGACACCGGTGTCACACGAGC----- 1054

QY 42 MetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrlleAlaLeuGlyArgThrAsnAsn 61
 Db 1055 ---CATCCGACCTGAAGGCGCGGTGACCGCTCCAAGAACTTCAACGCGCGCGCGC 1111
 QY 62 AlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlyAla 81
 Db 1112 GCGGCGCAAGGTGGGCCACGACCCAGCTCGCTCGATCGCGCGGCGGCGGCGGC 1171

QY 82 ThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIle 97
 Db 1172 CAGTCCCAAGGCGAAGTACAAAGGCGGTGCGACCGCGCGCGCGATCTCAACGCGAAGT 1231
 QY 98 MetAspSerSerGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGln 117
 Db 1232 CTCGACGACTCCGT-----TTCGCGGACGACTCCGCACTCTCGCGCGCATGGAGTGG 1285

QY 118 AlaPheSerAlaGlyAlaArgIle-HisThrAsnSerTrpGlyAlaAlaValAsnGlyAl 137
 Db 1286 GCG 1345
 QY 137 aTyThrThrAspSerArgAsnValAspAspTyrlleValArgLysAsnAspMetThrIleLe 157
 Db 1346 ACCGACCCGCTGGAGCGCGG-GTCGACAAAGCTGTTCGCGCGAAGCGCGCTCTGTTCGC 1404

QY 157 uPheAlaAlaGlyAsnGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLy 177
 Db 1405 CATCG 1458
 QY 177 sAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrlleAs 197
 Db 1459 CGCGCGCTCACCGCTCG 1485

QY 197 pAsnIleAsnHisValAlaGlnPheSerSerArgGlyPro---ThrLysAspGlyArgIle 216
 Db 1486 CGACAAGGACAAGCTCG 1545
 QY 216 eLysProAspValMetAlaProGlyThrTyrlleLeuSerAlaArgSerSerLeuAlaPr 236
 Db 1546 CAAGCGGAGGTCAACCGCTCCCGCGGTGGACATACGCGCGCGCGCGCGCGCGCGCGCGCA 1605

```

QY 236 oAspSerSerPheTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyThrSerMe 256
Db 1606 CATCGCCAGGAGTCCGGTGGAGGACCGCGCTACATGACCACTCCGGCAGCTCGAT 1665
QY 256 tAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnAr 276
Db 1666 GCGACCCCGCAGCTCGCGGGCGCGCCCTCTCTGAGAGCAGCAG----- 1711
QY 276 gGlyIleThrProLysProSerLeuLeuLysAlaLeu-----IleAlaGlyAl 293
Db 1712 -----CACCCCGACTGGACCTCCCGCGAAGTGAAGGGCGGCTCACCGGCTC 1758
QY 293 aAlaAspValGlyLeuGlyTyrProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAs 313
Db 1759 CACCAAGGCGGC---AAGTACACCCCGTTCGAGCAGGTTTCGGCCCGGATCCAGGCCGA 1815
QY 313 pLysSerLeuAsnValAlaTyrValAsnGluSerSerAlaLeuSer----- 328
Db 1816 CRAGGCGCTCCAGCAGACCGTGTACCGCAGCCGGTCTCGGTGAGCTTCGGGCTCAGCA 1875
QY 329 -----ThrSerGlnLysAlaThrTyrThrPheThrAlaTh 340
Db 1876 GTGCGCGCACACCGACGAGCGGTCCACCAAGCAGCTGACCTACCGCAACTCGGCAC 1935
QY 340 rAlaGlyLysProLeuLysIleSerLeuValTrpSerAsp-----AlaPr 355
Db 1936 CCAGGAGCTCAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 1995
QY 355 oAlaSer-----ThrThrAlaSerValThrLeuValAsnAspLeuAspLeuValIleTh 373
Db 1996 GCGCGGGCTTCTTACCGTGGCGCCACACG-----GTGAC 2031
QY 373 rAlaProAsnGlyThrArgTyrValGlyAsnAspPheSerAlaProPheAspAsnAsnTr 393
Db 2032 CTRCCCGCGGCGCGC-----AGCGCTCCGTGACATGACCGC 2070
QY 393 pAspGlyArgAsnAsnValGluAsnValPheIleAsnSerProGlnSerGlyThrTyrTh 413
Db 2071 CGACACCGG-----CTCGCGCGCAGCGTGGACGCGGTACTC 2109
QY 413 rIle----- 414
Db 2110 GCGTACGTGTGCCACCGCGCGCGCGGAGAGCGTCCGCGCGCGCGCGGTGCGCG 2169
QY 415 -GluValGlnAlaTyrAsnValProVal 423
Db 2170 CGAGTCTGAGTCTGACGAGTGTACCGTGC 2197

RESULT 15
AAT61455
ID AAT61455 standard; DNA; 2809 BP.
XX
AC AAT61455;
XX
DT 17-OCT-2003 (revised)
DT 06-OCT-1997 (first entry)
XX
DE DhpA-mel chimeric gene.
XX
KW asymmetric hydrolase; dhpA; 4-substituted-1,4-dihydropyridine;
KW derivative; Streptomyces viridosporus; ester; chiral; synthesis;
KW cardiovascular; treatment; hypertension; ischaemic heart disease; ds.
XX
OS Streptomyces viridosporus.
OS Streptomyces antibioticus.
OS Chimeric.
XX
FH Key Location/Qualifiers
CDS 338..2809
FT /*tag= a
FT 338..2539
FT mat_protein /*tag= b

```

```

FT
FT /product= "DhpA_protein_product"
FT /note= "from S. viridosporus"
FT mat_protein 2540..2809
FT /*tag= c
FT /product= "Melanin"
FT /note= "from S. antibioticus"
XX
XX WO9705243-A1.
XX
XX 13-FEB-1997.
XX
XX 30-JUL-1996; 96WO-JP002147.
XX
XX 31-JUL-1995; 95JP-00213975.
XX 29-FEB-1996; 96JP-00067478.
XX
XX (SAOC ) MERCIAN CORP.
XX
XX Arisawa A, Matsufuji M, Tsuruta T, Dobashi K, Nakashima T;
XX Isehiki K, Yoshioka T;
XX
XX WPI: 1997-145682/13.
XX P-PSDB; AAW13668.
XX
XX Asymmetric hydrolase gene derived from Streptomyces viridosporus - acts
XX on 4-substituted-1,4-di:hydro:pyridine derivatives to produce chiral
XX derivatives useful for synthesis of cardiovascular drugs.
XX
XX Claim 5; Page 37-43; 78pp; Japanese.
XX
XX This sequence is a fusion gene encoding Streptomyces viridosporus dhpA
XX gene product (an asymmetric hydrolase) which acts on 4-substituted-1,4-
XX dihydropyridine derivatives, and melanin from S. antibioticus. The dhpA
XX enzyme allows the efficient conversion of 4-substituted-1,4-
XX dihydropyridine esters to chiral partially hydrolysed derivatives, for
XX use in the synthesis of cardiovascular drugs suitable for the treatment
XX of e.g. hypertension and ischaemic heart disease. (Updated on 17-OCT-2003
XX to standardise OS field)
XX
XX Sequence 2809 BP; 473 A; 1041 C; 984 G; 311 T; 0 U; 0 Other;
XX

Alignment Scores:
Pred. No.: 1,27e-18 Length: 2809
Score: 354.50 Matches: 139
Percent Similarity: 41.49% Conservative: 56
Best Local Similarity: 29.57% Mismatches: 173
Query Match: 15.79% Indels: 103
DB: 2 Gaps: 17

US-09-985-689A-2 (1-434) x AAT61455 (1-2809)
QY 2 AspValAlaArgGlyIleValLysAlaAspValAlaGlnSerTyrGlyLeuTyrGly 21
Db 953 GACACGTCCGTCCGGCAGATCGGCGCCCGCCCAAGGGGTGTCGCC---GGCTACGCGGC 1009
QY 22 GlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 41
Db 1010 AAGGCGGTGAAGATCGCGTCTGTGACACCGCGTGTGACACGAGC----- 1054
QY 42 MetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsn 61
Db 1055 ---CATCCGACCTGAGGGCGGGGTGACCGCTCCCAAGAACCTCACCGCGCGCGCGC 1111
QY 62 AlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlyAla 81
Db 1112 GCGCGCGACAGGTGGCGCCACGCGCACCGCTCGCTCGATCGCGCGCGCGCACGCGCGCC 1171
QY 82 ThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIle 97
Db 1172 CAGTCCAAAGGGCAAGTACAAAGGGCGTCCACCGCGCGCGCGCTCAACGCGCAAGTTC 1231
QY 98 MetAspSerSerGlyGlyLeuGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGln 117

```

Search completed: April 4, 2004, 01:10:34
Job time : 365.684 secs

[illegible]

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 3, 2004, 23:50:44 ; Search time 63.5977 Seconds
(without alignments)
3787.066 Million cell updates/sec

Title: US-09-985-689a-2
Perfect score: 2245
Sequence: 1 NDVARGIVKADVAQSSVGLY.....EVQAYNPVGPQNFSLAIVN 434

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-O=/cgn2_1/USPTO spool/US09985689/runat 31032004 161807 4184/app query.fasta_1.3498
-DB=Issued Patents NA -QWTF=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pco -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09985689 @CGN_1.1.142 @runat 31032004 161807 4184 -NCFU=6 -ICPU=3
-NO MMAP -LARGQUERY -NEG SCORE=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:

1: /cgn2_6/ptodata/2/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS.COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length DB | ID | Description |
|------------|--------|-------------|-----------|----|-------------------|
| 1 | 2237 | 99.6 | 1920 | 4 | US-09-509-814A-3 |
| 2 | 2197 | 97.9 | 1923 | 4 | US-09-509-814A-7 |
| 3 | 2191 | 97.6 | 1923 | 4 | US-09-509-814A-5 |
| 4 | 2118.5 | 94.4 | 2003 | 2 | US-08-873-479-41 |
| 5 | 457.5 | 20.4 | 1977 | 3 | US-08-894-818B-2 |
| 6 | 457.5 | 20.4 | 1977 | 4 | US-09-445-472-11 |
| 7 | 425.5 | 19.0 | 1236 | 4 | US-09-445-472-2 |
| 8 | 425.5 | 19.0 | 1566 | 3 | US-08-894-818B-4 |
| 9 | 425.5 | 19.0 | 1962 | 3 | US-08-894-818B-34 |
| 10 | 425.5 | 19.0 | 1962 | 4 | US-09-445-472-15 |
| 11 | 414.5 | 18.5 | 1977 | 3 | US-08-894-818B-6 |
| 12 | 359.5 | 16.0 | 2539 | 3 | US-09-000-016-3 |

| | | | | | | |
|----|-------|------|------|---|--------------------|--------------------|
| 13 | 359.5 | 16.0 | 2539 | 4 | US-09-514-340-3 | Sequence 3, Appli |
| 14 | 354.5 | 15.8 | 2809 | 3 | US-09-000-016-1 | Sequence 1, Appli |
| 15 | 354.5 | 15.8 | 2809 | 4 | US-09-514-340-1 | Sequence 1, Appli |
| 16 | 311.5 | 13.9 | 1859 | 3 | US-08-894-818B-15 | Sequence 15, Appli |
| 17 | 307 | 13.7 | 2835 | 1 | US-08-750-532-2 | Sequence 2, Appli |
| 18 | 307 | 13.7 | 4765 | 1 | US-08-750-532-8 | Sequence 8, Appli |
| 19 | 307 | 13.7 | 4765 | 3 | US-08-894-818B-7 | Sequence 7, Appli |
| 20 | 307 | 13.7 | 4765 | 4 | US-09-445-472-5 | Sequence 5, Appli |
| 21 | 302.5 | 13.5 | 2532 | 1 | US-07-671-376C-4 | Sequence 7, Appli |
| 22 | 300 | 13.4 | 898 | 1 | US-08-750-532-7 | Sequence 7, Appli |
| 23 | 291.5 | 13.0 | 564 | 3 | US-08-750-532-11 | Sequence 11, Appli |
| 24 | 291.5 | 13.0 | 564 | 3 | US-08-894-818B-14 | Sequence 14, Appli |
| 25 | 273 | 12.2 | 1306 | 4 | US-09-966-921A-1 | Sequence 1, Appli |
| 26 | 273 | 12.2 | 1330 | 4 | US-09-966-921A-5 | Sequence 5, Appli |
| 27 | 260.5 | 11.6 | 1194 | 4 | US-09-328-352-3407 | Sequence 3407, Ap |
| 28 | 259 | 11.5 | 1628 | 4 | US-09-634-238-212 | Sequence 212, App |
| 29 | 256 | 11.4 | 840 | 1 | US-08-434-255-7 | Sequence 7, Appli |
| 30 | 256 | 11.4 | 840 | 1 | US-08-459-967-7 | Sequence 7, Appli |
| 31 | 256 | 11.4 | 840 | 1 | US-08-460-327-7 | Sequence 7, Appli |
| 32 | 256 | 11.4 | 840 | 1 | US-08-459-871-7 | Sequence 7, Appli |
| 33 | 256 | 11.4 | 840 | 3 | US-09-024-532-1 | Sequence 1, Appli |
| 34 | 256 | 11.4 | 840 | 4 | US-09-104-632A-1 | Sequence 1, Appli |
| 35 | 256 | 11.4 | 840 | 4 | US-09-019-532-1 | Sequence 1, Appli |
| 36 | 256 | 11.4 | 840 | 4 | US-09-417-359A-1 | Sequence 1, Appli |
| 37 | 256 | 11.4 | 840 | 4 | US-09-705-185-1 | Sequence 1, Appli |
| 38 | 256 | 11.4 | 840 | 4 | US-09-338-746-1 | Sequence 1, Appli |
| 39 | 256 | 11.4 | 1110 | 1 | US-08-434-255-5 | Sequence 5, Appli |
| 40 | 256 | 11.4 | 1110 | 1 | US-08-459-967-5 | Sequence 5, Appli |
| 41 | 256 | 11.4 | 1110 | 1 | US-08-460-327-5 | Sequence 5, Appli |
| 42 | 256 | 11.4 | 1110 | 1 | US-08-459-871-5 | Sequence 5, Appli |
| 43 | 256 | 11.4 | 1191 | 1 | US-08-434-255-3 | Sequence 3, Appli |
| 44 | 256 | 11.4 | 1191 | 1 | US-08-459-967-3 | Sequence 3, Appli |
| 45 | 256 | 11.4 | 1191 | 1 | US-08-460-327-3 | Sequence 3, Appli |

ALIGNMENTS

RESULT 1

US-09-509-814A-3
; Sequence 3, Application US/09509814A
; Patent No. 6376227
; GENERAL INFORMATION:
; APPLICANT: TAKAIWA, MIKIO
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SAEKI, KATSUHIKA
; APPLICANT: KUBOTA, HIROMI
; APPLICANT: HITOMI, JUN
; APPLICANT: KAGEYAMA, YASUSHI
; APPLICANT: SHIKATA, SHITSUM
; APPLICANT: NOMURA, MASAFUMI
; TITLE OF INVENTION: ALKALINE PROTEASE
; FILE REFERENCE: 0327-0832-0PCT
; CURRENT APPLICATION NUMBER: US/09/509,814A
; CURRENT FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: PCT/JP98/04528
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: JP 9-274570
; PRIOR FILING DATE: 1997-06-08
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 3
; LENGTH: 1920
; TYPE: DNA
; ORGANISM: Bacillus sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1920)
US-09-509-814A-3

Alignment Scores:
Pred. No.: 3 9e-224 Length: 1920
Score: 2237.00 Matches: 433
Percent Similarity: 99.77% Conservative: 0

Db 739 TCATGTCATGAAGCCCTCCGCGGGAATAATCTGCTATTATATGCTATTCGCGGACGACGAAT 798
Qy 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
Db 799 AATGCCAATGATACGAATGGTATGGTACGCGATGGCTGGCTCGTATTAGGAAACGGC 858
Qy 81 AlaThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
Db 859 TCACCTAATAAAGGAATGGCGCTCAGGCGAATCTAGTCTTCCTCAATCTATCATGGATAGC 918
Qy 101 SerGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaPheSer 120
Db 919 GGTGGGGACTTGGAGGACTCTTCGAATCTGCAACCTTATTCAGCCAGCATACTAGT 978
Qy 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaIleThrThr 140
Db 979 GGTGGTGGCAGAAATTCATCAAACTCCTGGGAGCAGCAGTGAATGGGGCTTACACA 1038
Qy 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
Db 1039 GATTCAGAAATGTGGATGACTATGTGGCAAAATGATATGACGATCTCTTTTGGCTGCC 1098
Qy 161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
Db 1099 GCGAATGAAGGACCGAACCAGCGGAACCATCAGTCGCCAGCAGGACAGCTAAATAATGCAATA 1158
Qy 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
Db 1159 ACAGTCGGAGCTACGGAAACCTCCGCCAACCTTTGGTCTTATGCCGACAATATCAAC 1218
Qy 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
Db 1219 CATGTGGCAGTCTCTTCAGTCGGACCGACAAAGGATGGACGATCAAAACCGGATGTC 1278
Qy 221 MetAlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
Db 1279 ATGGCACCGGGAACGTCTATCTATCAGCAAGATCTCTCTTCGACCGGATCTCCCTTC 1338
Qy 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260
Db 1339 TGGCGCAACCATGACAGTAAATATGATACATCATGGTGGACGCTCCATGGCTACACCGATC 1398
Qy 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
Db 1399 GTTCTGGAACGTGGCAGCTCTGTGAGCAATTTGTGAATAACAGAGGCATCACCA 1458
Qy 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspValGlyLeuGlyTyr 300
Db 1459 AGCCCTTCTCTATTAAAGCGGCACCTGATGGCGGTGACGCTGACATCGGCTTGGCTAC 1518
Qy 301 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
Db 1519 CGAACGGTAACCAAGGATGGGACGAGTGACATTTGGATAATCCCTGAAAGCTTGCCTAT 1578
Qy 321 ValAsnGluSerSerAlaLeuSerThrSerGlnLysAlaThrTyrThrPheThrAlaThr 340
Db 1579 GTGACAGCTCCAGTCTCTATCCACAGCCAAAGCGAGCTACTCGTTACTGCTACT 1638
Qy 341 AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrThrAla 360
Db 1639 GCGGCAAGCCTTGAATAATCTCCCTGGTATGGTCTGATGCCCTGCCAGCACACTGCT 1698
Qy 361 SerValThrLeuValAsnAspLeuPheValIleThrAlaProAsnGlyThrArgTyr 380
Db 1699 TCGTAACGCTTGTCAATGATCTGGACCTTGTCATTTACCGCTTCCAAATGGCACACAGTAT 1758
Qy 381 ValGlyAsnAspPheSerAlaProPheAsnAsnTrpAspGlyArgAsnAsnValGlu 400
Db 1759 GTAGGAATGACTTTACTTCGCCATACATGATACTGGATGGCGCGCAATAACGTAGAA 1818
Qy 401 AsnValPheIleAsnSerProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
Db 1819 AATGTAATTAATGACCAACAAAGCGGAGCTATACAAATGGAATGAGTACAGCTTATAAC 1878

Qy 421 ValProValGlyProGlnAsnPheSerLeuAlaIleValAsn 434
Db 1879 GTACCGTTGGACACAGAACTTCTCGTTGGCAATTGTGAAT 1920

RESULT 3

US-09-509-814A-5
; Sequence 5, Application US/09509814A
; Patent No. 6376227
; GENERAL INFORMATION:
; APPLICANT: TAKAIWA, MIKIO
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SAEKI, KATSUHIKA
; APPLICANT: KUBOTA, HIROMI
; APPLICANT: HITOMI, JUN
; APPLICANT: KAGEYAMA, YASUSHI
; APPLICANT: SHIKATA, SHITSUW
; APPLICANT: NOMURA, MASAFUMI
; TITLE OF INVENTION: ALKALINE PROTEASE
; FILE REFERENCE: 0327-0832-0PCT
; CURRENT APPLICATION NUMBER: US/09/509,814A
; CURRENT FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: PCT/JP98/04528
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: JP 9-274570
; PRIOR FILING DATE: 1997-06-08
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 5
; LENGTH: 1923
; TYPE: DNA
; ORGANISM: Bacillus sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1923)
US-09-509-814A-5

Alignment Scores:

Pred. No.: 2,54e-219 Length: 1923
Score: 2191.00 Matches: 419
Percent Similarity: 99.54% Conservative: 13
Best Local Similarity: 96.54% Mismatches: 2
Query Match: 97.59% Indels: 0
DB: 4 Gaps: 0

US-09-985-689A-2 (1-434) x US-09-509-814A-5 (1-1923)

Qy 1 AsnAspValAlaAlaArgGlyIleValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
Db 619 ATGATGTTCCGCGTGGAAATGTCAAAGCGGATGGCTCAGACGAGCTACGGGTGTAT 678
Qy 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 679 GGACAGGACAGATCGTAGCGGTTGCCGATACAGGGCTTGATACAGGTCCCAATGACAGT 738
Qy 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 739 TCATGTCATGAAGCCTTCCCGGGAAATTTACTGCAATTTATGCTATTCATTGGACGAGCAAT 798
Qy 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
Db 799 AATGCCAATGATACGAATGGTATGTCATGTCATGTCATGTCGCTGGCTCCGTATTAGGAACGGC 858
Qy 81 AlaThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
Db 859 TCACCTAATAAAGGAATGGCGCTCAGGCGAATCTAGTCTTCCAATCTATCATGGATAGC 918
Qy 101 SerGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaPheSer 120
Db 919 GGTGGGGACTTGGAGGACTACTCTTCGAATCTGCAAACTTATTCAGCCAGCATACTAGT 978
Qy 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyrThrThr 140

979 GCTGTGCCAGAAATTCATAAACTCCTGGGAGCAGCAGTGAATGGGGCTTACACAACA 1038
141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
1039 GATTCAGAAATGTGATGACATATGTCGCAAAATGATGACGATCTTTTCCTGCC 1098
161 GlyAsnGluGlyProAsnGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
1099 GGAATGAAGGACCGCAACCGGCAACCATCAGTCACCGGACGACGCTAAATAATCAATA 1158
181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
1159 ACAGTCGGAGCTACGGAAACCTCCGCCAAGCTTTGGTCTTATGCGGACATATCAAC 1218
201 HisValAlaGlnPheSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
1219 CATGTGCGCAGTCTCTTCACGTGACCGCACAAAGGATGGACGATCAACCGGATGTC 1278
221 MetAlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
1279 ATGGCACCAGGAAACGTTTACATATCAGCAAGATCTTCTTGCACCGGATTCCTCTTC 1338
241 TrpAlaAsnHisAspSerIstYrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260
1339 TGGCGCAACCATGACAGTAAATATGATATGATGCGGTGGAACTCCATGGCTACACCGATC 1398
261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
1399 GTTGTCTGGAACGTGGCAGCTTCGTGAGCATTTTGTGAAAACAGAGGCTACACCA 1458
281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspValGlyLeuGlyTyr 300
1459 AAGCCTTCTCTATTAAGGCGCACTGATTCGGGTGGCACTGACATCGGCTTGGCTAC 1518
301 ProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
1519 CCAACCGTAAACAGATGGGACGAGTGACATTTGGTAAATPCCCTGACGCTGCTAT 1578
321 ValAsnGluSerSerAlaLeuSerThrSerGlnLysAlaThrTyrThrPheThrAlaThr 340
1579 GTGAACGAGTCCAGTTCTCTATCCACAGCAAAAGCGAGCTACTCGTTTACTGTACT 1638
341 AlaGlyLysProLeuLysIleSerLeuValTyrSerAspAlaProIleSerThrAla 360
1639 GCGCGCAAGCCTTTGAAATCTCCCTGGTATGCTGTGATGCCCTCGGACCAACTGCT 1698
361 SerValThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrArgTyr 380
1699 TCCGTAACGCTTGTCAATGATCTGGACCTTGTCAATCCGCTCCAAATGGCACACAGTAT 1758
381 ValGlyAsnAspPheSerAlaProPheAspAsnAsnTrpAspGlyArgAsnAsnValGlu 400
1759 GTAGGAATGACVTTTACTTCGCCATACATGATGATGATGATGATGATGATGATGATGAT 1818
401 AsnValPheIleAsnSerProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
1819 AATGTATTTTATATGACACCAAAAGCGGACGCTATACATTTGAGGTACAGGCTTATAAC 1878
421 ValProValGlyProGlnAsnPheSerLeuAlaIleValAsn 434
1879 GTACCGGTGGACACAGACCTTCTGTTGGCAATTTGTAAT 1920

RESULT 4

US-08-873-479-41
; Sequence 41, Application US/08873479
; Patent No. 5891701
; GENERAL INFORMATION:
; APPLICANT: Sloma, Alan
; APPLICANT: Lynne, Christiantson
; TITLE OF INVENTION: Nucleic Acids Encoding A Polypeptide
; TITLE OF INVENTION: Having Protease Activity
; NUMBER OF SEQUENCES: 57
; CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 5891701o No. 5891701disk of No. 5891701th America
STREET: 405 Lexington Avenue
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10174
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/873,479
FILING DATE: 12-JUN-1997
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Agris, Cheryl H
REGISTRATION NUMBER: 34,086
REFERENCE/DOCKET NUMBER: 5251.000-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
TELEX:
INFORMATION FOR SEQ ID NO: 41:
SEQUENCE CHARACTERISTICS:
LENGTH: 3003 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
US-08-873-479-41
Alignment Scores:
Pred. No.: 1,98e-211 Length: 3003
Score: 2118.50 Matches: 405
Percent Similarity: 97.47% Conservative: 18
Best Local Similarity: 93.32% Mismatches: 10
Query Match: 94.37% Indels: 1
DB: 2 Gaps: 1
US-09-985-689A-2 (1-434) x US-08-873-479-41 (1-3003)
QY 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
DB 1470 AATGACGTGGCCCTGGCATTTGAAAGCAGAGTCGCAAAATACCTTTGGCTTATAT 1529
QY 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
DB 1530 GGACAAGCAGACAGATTGTAGCAGTTGCTGATCTGGGCTTGATACAGGAGAAATGACAGT 1589
QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
DB 1590 TCGATGATGATGAGCATTTCCGCGGTAGATTACCGCATATATGACCTATGCTGGGACAGCAAT 1649
QY 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
DB 1650 AACGCCAATGATCCAAATGGACATGGAAACCATGTTCTGCTGATCTGTGTAGGAAT --- 1706
QY 81 AlaThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
DB 1707 GCTCAAAATAAAGGATGGCAGCCCAATCTAGCTTTTCAATCTATTATGGATAGT 1766
QY 101 SerGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaPheSer 120
DB 1767 GGTGAGGCTGGGAGGACTACCTGCTAACTCAAAACATTTATTCAGTCAAGCATATAGT 1826
QY 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyrThr 140
DB 1827 GCTGAGCGAGAAATTCATCAAGATTCATGGGGGCTCCAGTAAACGGTGGCTTACGACA 1886
QY 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
DB 1887 GACTCTCGAAATGTTGATGATTATGTGAGAAAAAATGATATGACGATCTTTTTCGGGCC 1946

QY 161 GlyAsnGluGlyProAsnGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
Db 1947 GGAATAGGACCGACAGGAGCGTACATCAGTGCACCGAAGCAAAATATGCGATT 2006
QY 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerThrAlaAspAsnIleAsn 200
Db 2007 ACAGTTGGGCAACCGCAAACTACGTCACAGTTCCGATCTTATCGGATAATTAAC 2066
QY 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
Db 2067 CATGTTGCTCAATCTCTTCAGAGGCTCTACTAGATGACGCTATTAACCGCGAGTC 2126
QY 221 MetAlaProGlyThrIleLeuSerAlaArgSerLeuAlaProAspSerSerPhe 240
Db 2127 ATGGCACCAGGTACGTATATCTCTCTCTAGATCATCATAGTCCAGATTCCTCATTC 2186
QY 241 TrpAlaAsnHisAspSerLysThrAlaIleThrMetGlyThrSerMetAlaThrProIle 260
Db 2187 TGGGCAAAACCATGATAGTAATATCCCTACATGGTGGTGTACTTCTATGGCTTACTCCAAT 2246
QY 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
Db 2247 GTAGCAGGTATGTCACATTAAGGAGGATTTGTGAATAATAGGGGTAACTCCT 2306
QY 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspValGlyLeuGlyTyr 300
Db 2307 AAGCTTCCTCTTTAAAGCTGCTTTAATTCAGGTGCTGGGATGTTGGACTTGGCTTT 2366
QY 301 ProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
Db 2367 CCNATGTATACCAAGATGGGAGAGTACGTAGTATTAATCCCTAATGTGCAAT 2426
QY 321 ValAsnGluSerSerAlaLeuSerThrSerGlnLysAlaThrIleThrPheThrAlaThr 340
Db 2427 GTGAATGAACAGCGCTTTATCAACAGTCAAAAGCAACATATTCGTTACGGCTCAA 2486
QY 341 AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrAla 360
Db 2487 GCTGTAAACCCCTTAAATAATATCACTTGTGGTCAGATGCACCGGTAGCACCGCA 2546
QY 361 SerValThrLeuValAsnAspLeuValIleThrAlaProAsnGlyThrArgTyr 380
Db 2547 TCACTAACCTTTAGTGAATGATTTAGACTTATGATCACTGCACCAAAATGGAATATAC 2606
QY 381 ValGlyAsnAspPheSerAlaProPheAspAsnAsnTrpAspGlyArgAsnAsnValGlu 400
Db 2607 GTCGAAATGACTTTACAGCACCGTATGATACAAATGGGATGCGAAGAACACGTCGAA 2666
QY 401 AsnValPheIleAsnSerProGlnSerGlyThrIleGluValGlnAlaTyrAsn 420
Db 2667 AATGTTTATCAATGCTCTCAAGCGGAAGTATACAGTCAAGTGCAGGCTTACAT 2726
QY 421 ValProValGlyProGlnAsnPheSerLeuAlaIleValAsn 434
Db 2727 GTACCAAGTAAGTCGCAAAACCTTTCTTTAGGATTTGATAC 2768

RESULT 5

US-08-894-818B-2
; Sequence 2, Application US/08894818B
; Patent No. 6261822
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: YAMAMOTO, Katsuhiko
; APPLICANT: MITTA, Masanori
; APPLICANT: ASADA, Kiyozo
; APPLICANT: TSUNASAWA, Susumu
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Browdy and Neimark

STREET: 419 Seventh Street N.W., Ste. 300
CITY: Washington
STATE: D.C.
COUNTRY: United States of America
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/894,818B
FILING DATE: 20-MAY-1998
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP96/03253
FILING DATE: 07-NOV-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 323285/1995
FILING DATE: 12-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Browdy, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: TAKAKURA=1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-5197
TELEFAX: (202) 737-3528
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1977 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
US-08-894-818B-2

Alignment Scores:

Pred. No.: 5,998-38 Length: 1977
Score: 457.50 Matches: 136
Percent Similarity: 44.66% Conservative: 69
Best Local Similarity: 29.63% Mismatches: 153
Query Match: 20.38% Indels: 101
DB: 3 Gaps: 17

US-09-985-689A-2 (1-434) x US-08-894-818B-2 (1-1977)

QY 8 ValLysAlaAspValAlaGlnSerSerThrGlyLeuThrGlyGlnGlyIleValAla 27
Db 433 ATAGGGCCGATACCGTCTGGAACCTCCCTCGCTACGACGGAAGCGGTGTGGTGTGCC 492
QY 28 ValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisclualaPheArg 47
Db 493 ATCGTCGATACGGGTATAGACCGGAAC-----CACCCGATCTGAGG 534
QY 48 GlyLysIleThrAlaLeuTyr---AlaLeuGlyArgThrAsnAsnAlaAsnAspThrAsn 66
Db 535 GGCAAGGTCTATAGGTGTTAGCAGCCGCTCAACGCGAGGTGCGACCCCTTACGATGACCA 594
QY 67 GlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlyAlaThrAsnLys----- 84
Db 595 GGACACGGAACCCACGTTGCGGGTATCGTTGCGGAACCGCGGCTTAACCTCCAGTAC 654
QY 85 ---GlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerSer 101
Db 655 ATAGGCGTCCCGCCCGCGGCGAAGCTCGTCGCGGTCAAGTTCTCGGTGCCGACGTTCCG 714
QY 102 GlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaPheSerAla 121
Db 715 GGAAGCGTCTCCACCATCATCGCGGTGTGTGATCGGTGCTCCAGAACAGACAAAGTAC 774
QY 122 GlyAlaArgIle-----HisThrAsnSer 129
Db 775 GGATATAGGGTATCATCACTCTCTCCCTCGCTCTCCAGAGCTCCGACGGAACGACATCC 834

| | | | |
|----|------|--|------|
| Qy | 130 | TrpGlyAlaalaValAsnGlyAlaTyrThrAspSerArgAsnValAspAspTyrVal | 149 |
| Db | 835 | CTCAGTCAGCCCGTCAACAACGCGTGGAGCC | 867 |
| Qy | 150 | ArgLysAsnAspMetThrIleLeuPheAlaAlaGlyAsnGluGlyProAsnGlyGlyThr | 169 |
| Db | 868 | -----GGTATAGTAGTCTCGTCGCGCGCGCAACAGCGGCGCAACCTACACC | 918 |
| Qy | 170 | IleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArg | 189 |
| Db | 919 | GTCCGCTCACCCCGCGCGAGCAAGGTATACCGTCGGTGCA | 963 |
| Qy | 190 | ProSerPheGlySerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGly | 209 |
| Db | 964 | -----GTTGACAGCAACGACAACATCGCCAGCTTCTCCAGCAGGGA | 1005 |
| Qy | 210 | ProThrLysAspGlyArgIleLysProAspValMetAlaProGlyThrTyrIleLeuSer | 229 |
| Db | 1006 | CCAGCCGCGACGAAGGCTCAACCGGAAGTGTGCCCCCGCGGTGACATCATAGCC | 1065 |
| Qy | 230 | AlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsnHisAspSerLysTyrAla | 249 |
| Db | 1066 | CGCGCGCCAGC-----GGAACAGCATGGCGCCCGATTAACGACTACTACACC | 1116 |
| Qy | 250 | TyrMetGlyGlyThrSerMetAlaThrProIleValAlaGly---AsnValAlaGlnLeu | 268 |
| Db | 1117 | AAGCCCTCTGGAACACGATGGCCACCCCGACGTTTCGGCGGTGGCGGTCACTCCTC | 1176 |
| Qy | 269 | ArgGluHisPheValLysAsnArgGlyIleThrProLys-----ProSerLeuLeuLys | 286 |
| Db | 1177 | CAGCCCCAC-----CCGAGCTGGACCCCGGACAAGGTCAAG | 1212 |
| Qy | 287 | AlaAlaLeuIleAlaGlyAla-----AlaAspValGlyLeu | 298 |
| Db | 1213 | ACCGCCCTCATCGAGACGCGCGACATAGTCGCCCCAGAGAGATAGCGGCATCGCCTAC | 1272 |
| Qy | 299 | GlyTyrProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeu----- | 316 |
| Db | 1273 | GGTCG-----GGTAGGTGAACGCTCAAGGCCATCAAGTAC | 1311 |
| Qy | 317 | ---AsnValAlaTyrValAsnGluSerSerAlaLeuSerThrSerGlnLysAlaThrTyr | 335 |
| Db | 1312 | GACCACTACGCCAAGCTCATCTCCCGGCTCCGTGCGCGACAGGGAAGGCCACCCAC | 1371 |
| Qy | 336 | ThrPheThrAlaThrAlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaPro | 355 |
| Db | 1372 | ACCTTCGACGTACGGGGCGCCACTTCGTGACCGCCACCCCTCTACTGGAC----- | 1422 |
| Qy | 356 | AlaSerThrThrAlaSerValThrLeuValAsnAspLeuAlaIleThrAlaPro | 375 |
| Db | 1423 | -----ACGGGCTCGAGCGACATCGACCTCTACCTCTACGACCCC | 1461 |
| Qy | 376 | AsnGlyThrArgTyrValGlyAsnAspPheSerAlaProPheAspAsnAsnTrpAspGly | 395 |
| Db | 1462 | AACGGGAACGAG-----GTTGACTACTCTCTACACGCC | 1494 |
| Qy | 396 | ArgAsnAsnValGluAsnValPheIleAsnSerProGlnSerGlyThrTyrThrIleGlu | 415 |
| Db | 1495 | TACTACGGCTTCGAAAGGTGGCTACTACAAACCGCGCGGAACTGTGACGTCAAG | 1554 |
| Qy | 416 | ValGlnAlaTyrAsnValProValGlyProGlnAsnPheSerLeuAlaIleValAsn | 434 |
| Db | 1555 | GTGTCAGTCAACAG-----GGCGGGCGAACTACCAAGGTCAAGTGTGCTCAGC | 1602 |

RESULT 6

RESULT 6
US-09-445-472-11

US-09-443-472-11
; Sequence 11, Application US/09445472

; Sequence 11, Appendix 11
; Patent No. 6358726; FACE NO. 0358728
; GENERAL INFORMATION:

APPLICANT: TAKAKURA, Hikaru

APPLICANT: MORISHITA, Mio

APPLICANT: SHIMOJO, Tomoko

[illegible]

```

; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA=6
; CURRENT APPLICATION NUMBER: US/09/445,472
; CURRENT FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 11
; LENGTH: 1977
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-09-445-472-11

```

| Alignment Scores: | |
|------------------------|---------------|
| Pred. No.: | 5,99e-38 |
| Score: | 457.50 |
| Percent Similarity: | 44.66% |
| Best Local Similarity: | 29.63% |
| Query Match: | 20.38% |
| DB: | 4 |
| | DBs: |
| | Length: |
| | Matches: |
| | Conservative: |
| | Mismatches: |
| | Indels: |
| | Gaps: |
| | 1977 |

US-09-985-689A-2 (1-434) X US-09-445-472-11 (1-1977)

| | | | |
|----|-----|--|------|
| Qy | 8 | VallyAAlaspVallalaGlnSerSerTygLYeuTyrgLYcIndngLYnIleValAla | 27 |
| Dd | 433 | ATAGGGCCCATACCTCGTGAACCTCCTCGGTACTACGGAAGCGGTGGTGTTGCC | 492 |
| Qy | 28 | ValAlaspThrGLYLeuAspThrGiYAgaAsnAspSerSerMethIsGLuAlaPheArg | 47 |
| Dd | 493 | ATCGTCGATACGGGTATAGACGGCAAC-----CACCCGATCTGAAG | 534 |
| Qy | 48 | GlyLySileThraLaleuTyrr---AlaLeuGLYAgrThraSnAsnAlaasnAspThraSn | 66 |
| Dd | 535 | GGCAAGGTCTAGGCTGGTAGCGCCGCTCAAGCGCAGGTGCAGCCCCCTACGATGACCAG | 594 |
| Qy | 67 | GlyHisGLYThrHisValAlaGLYSerValLeuGLYAsgnGLYAlaThraSniLys--- | 84 |
| Dd | 595 | GGACACGGAAACCCACGTTGGCGGTATCGTTGCCGGAACCGCGACGCTTAACCTCCCAGTAC | 654 |
| Qy | 85 | --GlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerSer | 101 |
| Dd | 655 | ATAGGCGTCGCCCCCGCGCGAGAAGCTCGTCGGCGTCAAGGTTCTCGGTGCGCAGCGTTCG | 714 |
| Qy | 102 | GlyGLYLeuGLYLeuLeuProSerAsnLeuGLnThrLeuPheSerGlnAlaPheSerAla | 121 |
| Dd | 715 | GGAAAGCGTCTCCACCATCATCGCGGGGTGTGAATTGGGTCTGCCAGAACAGGACAAGTAC | 774 |
| Qy | 122 | GlyAlaArgile-----HISThrAnser | 129 |
| Dd | 775 | GGGATAAGGGTCATCAACCTCTCCCTCGGTCTCTCCGAGAGCTCCGACGGAACCGACTCC | 834 |
| Qy | 130 | TTrpGLYalaaLaALanGLYAlyAtTyThrThraSpSerArgAsnValAspTyrVal | 149 |
| Dd | 835 | CTCAGTCAGGCCGTCACACAGCGCTGGAGCGCC----- | 867 |
| Qy | 150 | ArgLyAsnAspMetThriLeuPheAlaAdGLYasnGLuGLYProAnGLYGLYThr | 169 |
| Dd | 868 | -----GGTATAGTAGTCTGCGTCGCCCGCGCAACAGCGGGCGGAACACTACACC | 918 |
| Qy | 170 | IleSerAlaProGLYThrAlaLYAsnAlaIleThrValGLYAlaThrGIuaSnLeuArg | 189 |
| Dd | 919 | GTCGGCTCACCCGCCCGCGGAGCAAGGTCTATAACCGTCGGTGCA----- | 963 |
| Qy | 190 | ProSerPheGLYSerTYrAlaAspAsnIleasnHisValaGlnPheSerSerArgGly | 209 |
| Dd | 964 | -----GTTGACGCAACGACACATCCGCCAGCTCTCTCAGCAGGGGA | 1005 |
| Qy | 210 | ProThrLYsAspGLYAgrILEYsProAspValMetAlaProGLYThrTYrIleLeuSer | 229 |

Db 1006 CGAGCCGGGAGGAGGCTCAAGCGGAAAGTCGTCGCCCGCGGCTTGACATCATAGCC 1065
Qy 230 AlaArgSerSerLeuAlaProAspSerPheTrpAlaAsnHisAspSerLysTrpAla 249
Db 1066 CGCGCGGCCAGC-----CGAACAGCATGGCCACCCCGGATAAACGATCTACACC 1116
Qy 250 TyrMetGlyGlyThrSerMetAlaThrProLeuValAlaGly---AsnValAlaGlnLeu 268
Db 1117 AAGGCTCTGGAACCAAGCATGCCACCCCGCACGTTTCGGCGTTGGCGCTCATCCTC 1176
Qy 269 ArgGluHisPheValLysAsnArgGlyLeuThrProLys-----ProSerLeuLeuLys 286
Db 1177 CAGGCCAC-----CCAGCTGGACCCCGGACCAAGGTGAAG 1212
Qy 287 AlaAlaLeuLeuAlaGlyAla-----AlaAspValGlyLeu 298
Db 1213 ACCGCCCTCATCGAGACCGCCGACATAGTCGCCCCCAAGGATAGCGCATCGCCTAC 1272
Qy 299 GlyTyrProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeu----- 316
Db 1273 GGTGGC-----GGTAGGGTGAACGCTCAAGGCCCATCAAGTAC 1311
Qy 317 ---AsnValAlaTrpValAsnGluSerSerAlaLeuSerThrSerGlnLysAlaThrTyr 335
Db 1312 GAGCACTACGCCAAGCTCACCTTCACCGCTCCGTCGCCGACCAAGGAGCGCCACCCAC 1371
Qy 336 ThrPheThrAlaThrAlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaPro 355
Db 1372 ACCTTCGACGTACGCGGCCACCTTCGTCGCCGCCACCCCTCTACTGGAC----- 1422
Qy 356 AlaSerThrThrAlaSerValThrLeuValAsnAspLeuAspLeuValIleThrAlaPro 375
Db 1423 -----ACGGCTCGAGCGACATCGACCTCTACTACGACCC 1461
Qy 376 AsnGlyThrArgTyrValGlyAsnAspPheSerAlaProPheAspAsnAsnTrpAspGly 395
Db 1462 AACGGGAACGAG-----GTTGACTCTCTACACGCC 1494
Qy 396 ArgAsnAsnValGluAsnValPheIleAsnSerProGlnSerGlyThrTyrThrIleGlu 415
Db 1495 TACTACGGCTTCGAGAGGTCCGGCTACTACACCGCGCGGCGGAACTGACCGGTCAAG 1554
Qy 416 ValGlnAlaTyrAsnValProValGlyProGlnAsnPheSerLeuAlaIleValAsn 434
Db 1555 GTCGTCAGCTACAAG-----GGCGCGGGAACCTACCAGGTCCAGCTCGTCAGC 1602

RESULT 7

/ Sequence 2, Application US/09445472
/ Patent No. 6358726
/ GENERAL INFORMATION:
/ APPLICANT: TAKAKURA, Hikaru
/ APPLICANT: MORISHITA, Mio
/ APPLICANT: SHIMOJO, Tomoko
/ APPLICANT: ASADA, Kiyozo
/ APPLICANT: KATO, Ikunoshin
/ TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
/ FILE REFERENCE: TAKAKURA=6
/ CURRENT APPLICATION NUMBER: US/09/445,472
/ CURRENT FILING DATE: 1999-12-06
/ PRIOR APPLICATION NUMBER: 151969/1997
/ PRIOR FILING DATE: 1997-06-10
/ NUMBER OF SEQ ID NOS: 33
/ SOFTWARE: PatentIn version 3.0
/ SEQ ID NO 2
/ LENGTH: 1236
/ TYPE: DNA
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Synthetic
US-09-445-472-2

Alignment Scores: 6.35e-35 Length: 1236
Pred. No.: 435.50 Matches: 1236
Score: 42.77% Conservative: 61
Best Local Similarity: 29.79% Mismatches: 147
Query Match: 18.95% Indels: 122
DB: 4 Gaps: 19

US-09-985-689A-2 (1-434) x US-09-445-472-2 (1-1236)

Qy 12 ValAlaGlnSerSerTyrGly-LeuTyr-----GlyGlnGly 24
Db 24 GTCTGCAGCTCACTTATGGCACTTACGTTGGAACTTGGGATATGATGTTCTGGAAT 83
Qy 24 nileValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisG1 44
Db 84 CACATATAGGAATAATTGCACACTGGAATTCAC-----GCTTCTCATCC 125
Qy 44 uAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAsn-- 63
Db 126 AGATCTCCAAAGGAAGAATA-----ATTGGTGGGTAGATTTGTCTCANTGG 170
Qy 64 -----AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuG1 78
Db 171 TAGGAGTTATCCATACGATGACCATGGACATGGAACCTCATGTAGCTTCAATAGCAGCTGG 230
Qy 78 yAsnGlyAlaThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPh 94
Db 231 TACTGAGCAGCAAGTAATGGCAAGTACAGGAATGGCTCCAGGAGTAAGCTGCGGG 290
Qy 94 eGlnSerIleMet-----AspSerSerGlyLeuGlyLeuGlyLeuProSerAsnLeuG1 112
Db 291 AATTAAGTTCTAGTGCCGATGGTTCTGGAACATATCTACTATAATTAAAGGAGTTGA 350
Qy 112 nThrLeuPheSerGlnAlaPheSerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAl 132
Db 351 GTGGCGGCTTGATACAAAGATAAGTACGGAATTAAGGTCAATTAATCTTCTTCTGTTTC 410
Qy 132 a-----AlaValAsnGlyValAlaTyrThr 140
Db 411 AAGCCAGAGCTCAGATGGTACTGACGCTCTAAGTCAGGCTGTTAATGACGCGTGGGATGC 470
Qy 140 rAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAl 160
Db 471 T-----GGATTAGTTGTTGTGGTTCGCCG 494
Qy 160 aGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaI1 180
Db 495 TGGAAACAGTGGACCTTAACAGTATACATCGTTCTCCAGCAGCTGCAAGCAAGTTAT 554
Qy 180 eThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAs 200
Db 555 TACAGTTGGAGCC-----GTTGCAAGATATGA 581
Qy 200 nHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVa 220
Db 582 TGTATTAACAGCTTCTCAGCAGAGGGCCAACTGCAGCGGCGGCTTAAGCTGAGGT 641
Qy 220 lMetAlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPh 240
Db 642 TGTGTCTCCAGGAACCTGGATAATTCTGCCAGAGCAAGT-----GGAACCTAGCAT 692
Qy 240 eTrpAlaAsnHisAspSerLysIleThrAlaTyrMetGlyGlyThrSerMetAlaThrPro1 260
Db 693 GGGTCAACCAATTAAATGACTATTACAGCAGCTCTCTGGGACATCAATGGCACTCTCTCA 752
Qy 260 eValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPr 280
Db 753 CGTAGCTGGTATTGACGCCCTCTTGTCCAA-----GCACACCC 791
Qy 280 oLys-----ProSerLeuLeuLysAlaLeuIleAlaGlyAla----- 293
Db 792 GAGCTGAGCTCCAGACAAAGTAAAAACAGCCCTCATAGAACTGCTGATATCGTAAAGCC 851

294 QY -----AlaaspValGlyLeuGlyTyrProAsnGlyAsnGlnGlyTrpDlyArgVa 310
 852 Db AGATGAATAGCCGATATAGCTACGGTGCA-----CGTAGGGT 890
 310 QY lThrLeuAspLysSerLeuAsnValAlaTyrValAsnGluSerSerAlaLeuSerThrSe 330
 891 Db TAATGTCATCAAGCGCTATAAC-----TAGGATACTATGCAAGCTAGTGTTCACCTGG 944
 330 QY rGlnLysAla-----ThrTyrThrPheThrAlaThrAlaGlyLysProLe 345
 945 Db ATATGTTGCCAACCAAGCAGCCAACTCACCAGTTCGTTATTAGCGGAGCTTCGTTCTGT 1004
 345 QY uLysIleSerLeuValTrpSerAspAlaProAsnSerThrThrAlaSerValThrLeuVa 365
 1005 Db AACTGCCACCATTTACTGGGACATGCGCAAT----- 1035
 365 QY lAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrArgTyrValGlyAsnAspPh 385
 1036 Db AGGACCTTGTAUCTTTACCTCTACGATCCCATGGAACAG----- 1077
 385 QY eSerAlaProPheAspAsnAsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAs 405
 1078 Db GTTGACTACTCTTACACCGCCTACTATGCGATTGCAAAAGGTGGTTATTATTA 1127
 405 QY nSerProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsnValProValGlyPr 425
 1128 Db CAACCCAACTGATGACATGCAATTAAGTTGTAAGCTACAGC-----GGAAG 1178
 425 QY oGlnAsnPheSerLeuAlaIleValAsn 434
 1179 Db TGCAAACTATCAAGTAGATGTTGGTAAAGT 1206

RESULT 8
 US-08-894-818B-4
 ; Sequence 4, Application US/08894818B
 ; Patent No. 6261822
 ; GENERAL INFORMATION:
 ; APPLICANT: TAKAKURA, Hikaru
 ; APPLICANT: MORISHITA, Mio
 ; APPLICANT: YAMAMOTO, Katsuhiko
 ; APPLICANT: MITTA, Masanori
 ; APPLICANT: ASADA, Kiyozo
 ; APPLICANT: TSUNASAWA, Susumu
 ; APPLICANT: KATO, Ikunoshin
 ; TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
 ; NUMBER OF SEQUENCES: 42
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Browdy and Neimark
 ; STREET: 419 Seventh Street N.W., Ste. 300
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: United States of America
 ; ZIP: 20004
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/894,818B
 ; FILING DATE: 20-MAY-1998
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/JP96/03253
 ; FILING DATE: 07-NOV-1996
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: JP 323285/1995
 ; FILING DATE: 12-DEC-1995
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Browdy, Roger L.
 ; REGISTRATION NUMBER: 25,618
 ; REFERENCE/DOCKET NUMBER: TAKAKURA=1

```
Db 642 TGTGCTCCAGAAACGGATAATGTGCGCAGACGAAT-----GGAACCTAGCAT 692
Qy 240 eTIPAlaasnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIl 260
Db 593 GGGTCAACCAATTAATGACTATTACACAGCAGCTCTGGACATCAATGCACTCTCTCA 752
Qy 260 eValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyLeThrPr 280
Db 753 CGTAGCTGGTATTGACGGCCCTCTGCTCCAA-----GCACACCC 791
Qy 280 oLys-----ProSerLeuLeuLysAlaAlaLeuAlaGlyAla----- 293
Db 792 GAGCTGAGCTCCAGACAAAGTAAACAGCCCTCTATAGAACTCTGATATCGTAAAGCC 851
Qy 294 -----AlaAspValGlyLeuGlyTyrProAsnGlyAsnGlnGlyTrpGlyArgVa 310
Db 852 AGATGAATACCGATATAGCTACCGGTGCA-----GGTAGGGT 890
Qy 310 lThrLeuAspLysSerLeuAsnValAlaTyrValAsnGluSerSerAlaLeuSerThrSe 330
Db 891 TAATGATACAGGCTATAAAC-----TAGATACTATGCAAGCTAGTGTTCCTGG 944
Qy 330 rGlnLysAla-----ThrTyrThrPheThrAlaThrAlaGlyLysProLe 345
Db 945 ATATGTTCCCAACAAAGGCAGCCAACTCACCAGTTGTTATTAGCGGAGCTTCGTTCTG 1004
Qy 345 uLysIleSerLeuValTrpSerAspAlaProAlaSerThrThrAlaSerValThrLeuVa 365
Db 1005 AACTGCCACATATATCTGGGCAATGCCAAT----- 1035
Qy 365 lAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrArgTyrValGlyAsnAspPh 385
Db 1036 -AGCGACCTTGATCTTACCTCTAGATCCCAATGGAAACCAAG----- 1077
Qy 385 eSerAlaProPheAspAsnAsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAs 405
Db 1078 -----GTTGACTACTCTTACACCCCTACTATGATTCGAAAGGTGTGTTATTA 1127
Qy 405 nSerProGlnSerGlyTyrTyrThrIleGluValGlnAlaTyrAsnValProValGlyPr 425
Db 1128 CAACCAACTGATGAAATGACATGACATTAAGTTGATAGCTACAGC-----GGAAG 1178
Qy 425 oGlnAsnPheSerLeuAlaIleValAsn 434
Db 1179 TGCAAACTCAAGTAGATGTGTAAGT 1206
```

RESULT 9

US-08-894-818B-34
Sequence 34, Application US/08894818B
Patent No. 6261822

GENERAL INFORMATION:

APPLICANT: TAKAKURA, Hikaru
APPLICANT: MORISHITA, Mio
APPLICANT: YAMAMOTO, Katsuhiko
APPLICANT: MITA, Masanori
APPLICANT: ASADA, Kiyozo
APPLICANT: TSUNASAWA, Susumu
APPLICANT: KATO, Ikunoshin
TITLE OF INVENTION: HYPER-THERMOSTABLE PROTEASE GENES
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:

ADDRESSEE: Browdy and Neimark
STREET: 419 Seventh Street N.W., Ste. 300
CITY: Washington
STATE: D.C.
COUNTRY: United States of America
ZIP: 20004

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/894,818B
FILING DATE: 20-MAY-1998
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP96/03253
FILING DATE: 07-NOV-1996
PRIOR APPLICATION DATA: JP 323285/1995
FILING DATE: 12-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Browdy, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: TAKAKURA-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-5197
TELEFAX: (202) 737-3528
INFORMATION FOR SEQ ID NO: 34:
SEQUENCE CHARACTERISTICS:
LENGTH: 1962 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
US-08-894-818B-34

Alignment Scores:

| Pred. No.: | 1-32e-34 | Length: | 1962 |
|------------------------|----------|---------------|------|
| Score: | 425.50 | Matches: | 140 |
| Percent Similarity: | 42.77% | Conservative: | 61 |
| Best Local Similarity: | 29.79% | Mismatches: | 147 |
| Query Match: | 18.95% | Indels: | 122 |
| DB: | 3 | Gaps: | 19 |

US-09-985-689A-2 (1-434) x US-08-894-818B-34 (1-1962)

| | | | |
|----|-----|--|----------------|
| Qy | 12 | ValAlaGlnSerSerTyrGly-LeuTyr----- | GlyGlnGlyCl 24 |
| Db | 420 | GTCTGAGCTCAAGTTATGCGCACTTACGTTGGAACTGGGATATGATGGTCTGGAAAT 479 | |
| Qy | 24 | nIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGl 44 | |
| Db | 480 | CACAAATAGGAATAATTGACACTGGAATTGAC-----GCTTCTCATCC 521 | |
| Qy | 44 | uAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAsn-- 53 | |
| Db | 522 | AGATCTCCAAGGAAAAAGTA-----ATTGGTGGGTAGATTTTGTCAATCG 566 | |
| Qy | 64 | -----AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGl 78 | |
| Db | 567 | TAGGAGTTATCCATACGATGACCATGGACATGGAACTCATGTAGCTTCAATAGCAGCTG 626 | |
| Qy | 78 | YAsnGlyAlaThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPh 94 | |
| Db | 627 | TACTGGAGCAGCAAGTAATGGCAAGTACAAGGAATGGCTCCAGGAGCTAAGCTGGCGGG 686 | |
| Qy | 94 | eGlnSerIleMet-----AspSerSerGlyGlyLeuGlyGlyLeuProSerAsnLeuGl 112 | |
| Db | 687 | AATTAAGGTTCAGTGGCCGATGGTCTCGAAGCATATCTACTATATTAAGGAGTTGA 746 | |
| Qy | 112 | nThrLeuPheSerGlnAlaPheSerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAl 132 | |
| Db | 747 | GTGGCGCGTTGATAACAAAGATAAGTACGGAATTAAGTCAATTAATCTTCTCTGTTTC 806 | |
| Qy | 132 | a-----AlaValAsnGlyValaTyrThrTh 140 | |
| Db | 807 | AAGCCAGAGCTCAGATGGTACTGACCGCTCTAAGCTAGGCTGTTTAAATGCGCGGTGGATGC 866 | |
| Qy | 140 | rAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAla 160 | |
| Db | 867 | T-----GGATTAGTTGTTGTTGTTGCGGC 890 | |
| Qy | 160 | aGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIl 180 | |

891 TGAACAGTGGACCTTAACAGTATACATCGTTCTCCAGCAGCTGCAAGCAAGTTAT 950
180 eThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerThrAlaAspAsnIleAs 200
951 TACAGTTGGAGCC-----GTTGACAAAGTATGA 977
200 nHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVa 220
978 TGTATTAACAAGCTTCTCAAGCAGAGGGCCAACTGCAGACGGCAGCTTAAGCTGAGGT 1037
220 lMetAlaProGlyThrThrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPh 240
1038 TGTGTCTCCAGGAACCTGATATGCTGCCAGACCAAGT-----GGAACATAGCAT 1088
240 eTrpAlaAsnHisAspSerLysThrAlaThrMetGlyGlyThrSerMetAlaThrProIl 260
1089 GSGTCACCAATTAATGACTATTACACAGCAGCTCTGGACATCAATGGCACTCTCA 1148
260 eValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPr 280
1149 CGTAGCTGTATTGACGCCCTCTTGTCTCAA-----GCACACCC 1187
280 oLys-----ProSerLeuLeuLysAlaAlaLeuIleAlaGlyAla----- 293
1188 GAGCTGGACTCCAGACAAAGTAAACACAGCCCTCATAGAACTGCTGATCGTAAAGCC 1247
294 -----AlaAspValGlyLeuGlyThrProAsnGlyAsnGlnGlyTrpGlyArgVa 310
1248 AGATGAATAGCGATATAGCTTACGTCGA-----GGTAGGGT 1286
310 lThrLeuAspLysSerLeuAsnValAlaThrValAsnGluSerSerAlaLeuSerThrSe 330
1287 TAATGATCAACAGCTATAAAC-----TACGATAACTATGCAAGCTAGTCTTCACTGG 1340
330 rGlnLysAla-----ThrTyThrPheThrAlaThrAlaGlyLysProbe 345
1341 ATATGTTGCCAACAAAGCGACCAAACTCACAGTTTCGTTATTAGCGGAGCTTCGTTCT 1400
345 uLysIleSerLeuValTrpSerAspAlaProAlaSerThrThrAlaSerValThrLeuVa 365
1401 AACTGCCACATTATCTAGGACATGCCAAT----- 1431
365 lAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrArgTyValGlyAsnAspPh 385
1432 -AGCGACCTTGACTTTACCTCTACGATCCCAATGGAACCCAG----- 1473
385 eSerAlaProPheAspAsnAsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAs 405
1474 -----GTTGACTACTCTTACACCGCCTACTATGGATTCCGAAAGGTTGGTTATTA 1523
405 nSerProGlnSerGlyThrTyThrIleGluValGlnAlaThrAsnValProValGlyPr 425
1524 CAACCCCACTGATGGACATGGACATTAAGTTGTAAGCTACAGC-----CGAAG 1574
425 oGlnAsnPheSerLeuAlaIleValAsn 434
1575 TGCAAACTATCAAGTAGATGTGGTAAGT 1602

RESULT 10

US-09-445-472-15
; Sequence 15, Application US/09445472
; Patent No. 6358726
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozi
; APPLICANT: KATO, Ikunoskin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA-6
; CURRENT APPLICATION NUMBER: US/09/445,472
; CURRENT FILING DATE: 1999-12-06

; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 15
; LENGTH: 1962
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-09-445-472-15

Alignment Scores: Length: 1962
Pred. No.: 132e-34 Matches: 140
Score: 425.50 Conservative: 61
Percent Similarity: 72.77% Mismatches: 147
Best Local Similarity: 29.79% Indels: 122
Query Match: 18.95% Gaps: 19
DB:

US-09-985-689A-2 (1-434) x US-09-445-472-15 (1-1962)

Qy 12 ValAlaGlnSerSerTyGly-LeuTy-----GlyGlnGlyGl 24
Db 420 GTCTGCAGCTCAAGTTATGCAACTTACGTTTGGAACTTGGGATATGATGTTCTCGAAT 479
Qy 24 nIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGl 44
Db 480 CACAAATAGGAATATTGACACTGGAATTGAC-----GCTTCTCATCC 521
Qy 44 uAlaPheArgGlyLysIleThrAlaLeuTyAlaLeuGlyArgThrAsnAsnAlaAsn-- 63
Db 522 AGATCTCCAAAGGAAAGTA-----ATTGGTGGGTAGATTTTCTCAATGG 566
Qy 64 -----AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGl 78
Db 567 TAGGAGTTATCCATACCATGACCATGACATGGAACATCATGTAGCTTCAATAGCAGCTGG 626
Qy 78 YAsnGlyValAlaThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPh 94
Db 627 TACTGGAGCAGCAAGTAATGCGAAGTACAAAGGAAATGGCTCCAGGAGCTAAGCTGGCGG 686
Qy 94 eGlnSerIleMet-----AspSerSerGlyGlyLeuGlyGlyLeuProSerAsnLeuGl 112
Db 687 AATTAAGGTTCTAGTGCCGATGTTCTTGAAGCATATCTACTATATTAATTAAGGAGTTGA 746
Qy 112 nThrLeuPheSerGlnAlaPheSerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAl 132
Db 747 GTGGCCCTTGCATTAACAAAGATAAGTACGGAATTAAGTTCATTAACTTTCTCTTGGTTC 806
Qy 132 a-----AlaValAsnGlyAlaTyThrTh 140
Db 807 AAGCCAGAGCTCAGATGCTACTGACGCTTAAGTCAGGCTGTTAATCGCGTGGGATGC 866
Qy 140 rAspSerArgAsnValAspAspTyValArgLysAsnAspMetThrIleLeuPheAlaAl 160
Db 867 T-----GGATTAGTTCTTGTGGTGGCGC 890
Qy 160 aGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIl 180
Db 891 TCGAAACAGTGGACCTTAACAAGTATACAATCGGTTCTCCAGCAGCTCCAGCAAGTTAT 950
Qy 180 eThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyAlaAspAsnIleAs 200
Db 951 TACAGTTGGAGCC-----GTTGACAAAGTATGA 977
Qy 200 nHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVa 220
Db 978 TGTATTAACAAGCTTCTCAAGCAGAGGGCCAACTGCAGACGGCAGCTTAAGCTGAGGT 1037
Qy 220 lMetAlaProGlyThrTyThrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPh 240
Db 1038 TGTGTCTCCAGGAACCTGATATGCTGCCAGACCAAGT-----GGAACATAGCAT 1088


```
QY 240 eTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIl 260
Db 1089 GGGTCAACCAATTAATGACTATTACACAGCAGCTCTGGGACATCAATGGCACTCCTCA 1148
QY 260 eValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPr 280
Db 1149 CGTAGCTGGTATGCAGCCCTCTGCTCCAA-----GCACACCC 1187
QY 280 oLys-----ProSerLeuLeuLysAlaAlaLeuIleAlaGlyAla----- 293
Db 1188 GAGCTGGAGCTCCAGACAAATAAACAGCCCTCATAGAACTGCTGATATCGTAAAGCC 1247
QY 294 -----AlaAspValGlyLeuGlyTyrProAsnGlyAsnGlnGlyTyrGlyArgVa 310
Db 1248 AGATGAATAGCGGATATAGCTACGGTGCA-----GATAGGGT 1286
QY 310 lThrLeuAspLysSerLeuAsnValAlaTyrValAsnGluSerSerAlaLeuSerThrSe 330
Db 1287 TAATGCATACAAGGCTATAAAC-----TACGATAACTATGCAAGCTAGTGTTCACCTGG 1340
QY 330 rGlnLysAla-----ThrTyrThrPheThrAlaThrAlaGlyLysProLe 345
Db 1341 ATATGTTGCCAACAAAGCAGCAACCTCAGGTTCTGTTATAGCGGAGCTTCGTTCTGT 1400
QY 345 ulysIleSerLeuValTyrPheAspAlaProAlaSerThrThrAlaSerValThrLeuVa 365
Db 1401 AACTGCCACATTATATCGGACAAATGCCAAT----- 1431
QY 365 lAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrArgTyrValGlyAsnAspPh 385
Db 1432 -AGCCACCTTGATCTTTCCTCTACGATCCCAATGGAACACAG----- 1473
QY 385 eSerAlaProPheAspAsnAsnTrpAspGlyArgAsnValGluAsnValPheIleAs 405
Db 1474 -----GTTGACTACTCTTACACCGCCTACTATCGATTGCAAGGTTGTTATTA 1523
QY 405 nSerProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsnValProValGlyPr 425
Db 1524 CAACCCAACTGATGGAACATGACAAATTAAGTTGTGTAAGCTACAGC-----GGAAG 1574
QY 425 oGlnAsnPheSerLeuAlaIleValAsn 434
Db 1575 TCACAACTATCAAGTAGATGTGGTAAGT 1602
```

RESULT 11

```
US-08-894-818B-6
Sequence 6, Application US/08894818B
Patent No. 6261822
GENERAL INFORMATION:
APPLICANT: TAKAKURA, Hikaru
APPLICANT: MORISHITA, Mio
APPLICANT: YAMAMOTO, Katsuhiko
APPLICANT: MITTA, Masanori
APPLICANT: ASADA, Kiyozo
APPLICANT: TSUNASAWA, Susumu
APPLICANT: KATO, Ikunoshin
TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESS: Browdy and Neimark
STREET: 419 Seventh Street N.W., Ste. 300
CITY: Washington
STATE: D.C.
COUNTRY: United States of America
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/894,818B
```

```
FILING DATE: 20-MAY-1998
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP96/03253
FILING DATE: 07-NOV-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 323285/1995
FILING DATE: 12-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Browdy, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: TAKAKURA=1
TELEPHONE: (202) 628-5197
TELEFAX: (202) 737-3528
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1977 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
US-08-894-818B-6
```

```
Alignment Scores:
Pred. No.: 1,89e-33 Length: 1977
Score: 414.50 Matches: 134
Percent Similarity: 43.07% Conservative: 68
Best Local Similarity: 28.57% Mismatches: 147
Query Match: 18.46% Indels: 120
DB: 3 Gaps: 19
```

US-09-985-689A-2 (1-434) x US-08-894-818B-6 (1-1977)

```
QY 12 ValAlaGlnSerSerTyrGly-LeuTyr-----GlyGlnGlyGl 24
Db 420 GTCTGCAGCTCAAGTATGCGCACTTACGTTTGGAACTTGGATATGATGTTCTGGAAT 479
QY 24 nIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGl 44
Db 480 CACATAGGAATAAATGACACTGGAATTGAC-----GCTTCTCATCC 521
QY 44 uAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAsn-- 63
Db 522 AGATCTCCCAAGGAAAGTA-----ATTGGTGGGTAGATTTTGTCAATGG 566
QY 64 -----AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGl 78
Db 567 TAGGAGTTATCCATACGATGACCATGCACTGCACTCATGTAGCTTCAATAGCAGCTGG 626
QY 78 yAsnGlyAlaThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPh 94
Db 627 TACTGGAGCAGCAAGTAATGGCAAGTACCAAGGAATGGCTCCAGGAGCTAAGCTGGCGG 686
QY 94 eGlnSerIleMet-----AspSerSerGlyGlyLeuGlyGlyLeuProSerAsnLeuGl 112
Db 687 AATTAAGGTTCTAGGTGCGCATGGTTCTTGGAAAGCATATCTACTATAATTAAAGGAGTTGA 746
QY 112 nThrLeuPheSerGlnAlaPheSerAlaGlyAlaArgIle----- 125
Db 747 GTGGGCGGTTGATACAAAGATAAGTACGGAATTAGGTCATTAAATCTTCTTCTTGGTTC 806
QY 126 -----HisThrAsnSerTrpGlyAlaAlaValAsnGlyValatyrThrTh 140
Db 807 AAGCCAGAGCTCCGACGGAACCGACTCCCTCAGTCAGCGCGCTCAACACGCTGGACGC 866
QY 140 rAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAl 160
Db 867 C-----GGTATAGTAGTCTCGTGGCGCGC 890
QY 160 aGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIl 180
Db 891 CGGCAACAGCGGCGCGGACACCTACACCTCGCTCGCTCACCGCGCGCGCGGCAAGGTCTAT 950
```

QY 180 eThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAs 200
Db 951 AACCGTGGTGCA-----GTTGACAGCAAGA 977
QY 200 nHisValAlaGlnPheSerArgGlyProThrIlyAspGlyArgIleIlyProAspVa 220
Db 978 CAACATCGCCAGCTTCTCAGAGGGGACCGCCGCGGAGAGGCTCAAGCGGAGT 1037
QY 220 iMetAlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPh 240
Db 1038 CTGCGCCCGGGGTGACATCATAGCCCGCGCCGAGC-----GGAACACGAT 1088
QY 240 eTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIl 260
Db 1089 GGGACCCCGGATAAAGACTACTACCAAGCCCTCTGGAACCAAGCATGGCCACCCGCA 1148
QY 260 eValAlaGlyIleAsnValAlaGlnLeuArgIleHisPheValIlyAsnAArgIlyIleTh 279
Db 1149 CGTTTCGGGGCGTGGCGGCTCATCTCCAGCCAC-----1185
QY 279 rProIlys-----ProSerLeuLeuIlyAlaAlaLeuIleAlaGlyAla-----293
Db 1186 -CCGAGCTGACCCCGGCAAGAGTGAAGCCGCTCATCGAGACCGCCGACATAGTCGC 1244
QY 294 -----AlaAspValGlyLeuGlyTyrProAsnGlyAsnGlnGlyTyrGlyAr 309
Db 1245 CCCAAGGAGATAGCGGACATCGCTACGTGG-----GGTAG 1283
QY 309 gValThrLeuAspLysSerLeu-----AsnValAlaTyrValAsnGluSerSerAl 326
Db 1284 GGTGAACGTCTACAGGCCATCAAGTACGACGACTACGCCAAGCTCACCTTCCCGGCTC 1343
QY 326 aLeuSerThrSerGlnIlyAlaThrTyrThrPheThrAlaThrAlaGlyLysProLeuIly 346
Db 1344 CGTCGGCGCAAGGAGCGCCACCCACCTTCGACGTCAGCGGCGCACCTTCGTGAC 1403
QY 346 sIleSerLeuValTrpSerAspAlaProAlaSerThrThrAlaSerValThrLeuValAs 366
Db 1404 CGCCACCTCTACTGGAC-----ACGGGCTCGAG 1433
QY 366 nAspLeuAspLeuValIleThrAlaProAsnGlyThrArgTyrValGlyAsnAspPheSe 386
Db 1434 CGACATCGACCTCTACTCTACGCCCAAGGAGGAGGAG-----1473
QY 386 xAlaProPheAspAsnAsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAsnSe 406
Db 1474 -----GTTGACTACTCTACACCGCTTACGAGGCTCGAGAGGCTCGGTACTACAA 1526
QY 406 rProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsnValProValGlyProCl 426
Db 1527 CCCGACCGCGGAACCTGGAGCTCAAGGTCTGCTACGCTACAG-----GGCGCGGC 1577
QY 426 nAsnPheSerLeuAlaIleValAsn 434
Db 1578 GAACCTACAGGTCTGACGCTCTGACG 1602

RESULT 12
US-09-000-016-3
; Sequence 3, Application US/09000016
; Patent No. 6143541
; GENERAL INFORMATION:
; APPLICANT: Akira ARISAWA et al.
; TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC
; TITLE OF INVENTION: HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-DIHYDROPYRIDINE DERIV
; TITLE OF INVENTION: ITS EXPRESSION PRODUCT
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
; STREET: 2033 K Street, N.W., #800
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.

ZIP: 20006
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/000,016
FILING DATE: January 30, 1998
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Warren M. Cheek, Jr.
REGISTRATION NUMBER: 33,367
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-721-8200
TELEFAX: 202-721-8250
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2539 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
ORIGINAL SOURCE:
ORGANISM: Streptomyces viridosporus
STRAIN: A-914
FEATURE:
NAME/KEY: CDS
LOCATION: 338...2539
IDENTIFICATION METHOD: E
US-09-000-016-3

Alignment Scores:
Pred. No.: 1.59e-27 Length: 2539
Score: 359.50 Matches: 141
Percent Similarity: 42.07% Conservative: 58
Best Local Similarity: 29.81% Mismatches: 165
Query Match: 16.01% Indels: 109
Gaps: 20

US-09-985-689A-2 (1-434) x US-09-000-016-3 (1-2539)

QY 2 AspValAlaArgGlyIleValIlyAspValAlaGlnSerSerTyrGlyLeuTyrGly 21
Db 953 GACACGTCGTCGGCAGATCGGCGCCCAAGCGGTGTCGCC-----GGCTACGACGCG 1009
QY 22 GlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSer 41
Db 1010 AAGGCGTGAAGATCGCGCTCTGACACCGGTGTCGACACGAGC-----1054
QY 42 MethisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsn 61
Db 1055 ---CATCCGACCTGAAGGGCGGTGACCGCTCAAGAACTTACCGCGCGCGCGC 1111
QY 62 AlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlyAla 81
Db 1112 GCCGCGACACAGGTGGGCGCACCGCACCGCTCGATCGCGGGCGGCGCGCGC 1171
QY 82 ThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIle 97
Db 1172 CAGTCCAGGCAAGTACAGGGGTGCGACCCCGCGCGCGCTCTCAACGGCAAGGTC 1231
QY 98 MetAspSerSerGlyLeuGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGln 117
Db 1232 CTCGACGACTCCGGT-----TTCCGGGACGACTCCCGCATCTCCGCGCATGGAGTGG 1285
QY 118 AlaPheSerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAla 137

| | | | | | |
|---|---|------|-------|--|------|
| D | b | 1286 | GC | CGCGCGCGAGCGGCGCGCGCTGCTCACCATGAGCGCTGGGC-----GGCATG | 1333 |
| Q | y | 138 | Ty | ThrThrAspSerArgAsnValAspAspTyValArgLys-----AsnAspMetThr | 155 |
| D | b | 1334 | GAC | ACCGGAGACACCGACCGCGTGGAGCGCGCGTGCACAAGCTGTCGCCGAGAGGCG | 1393 |
| Q | y | 156 | Ile | LeuPhe-----AlaAlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaPro | 173 |
| D | b | 1394 | GT | CTGTTCGCATCGCGCGCGCAACGAGGCGCCGGAG-----TCGATCGGTTCGCC | 1447 |
| Q | y | 174 | Gly | ThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGly | 193 |
| D | b | 1448 | GG | CAGCGCGAGCGCGCGCGCTCACCGTCGGCGCC----- | 1480 |
| Q | y | 194 | Ser | TyrAlaAlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGlyPro---ThrLys | 212 |
| D | b | 1481 | ----- | GTCCAGCAGCAAGACAAGCTCGCGCACTTCTCTCCACCGCGCGCGCTCGCG | 1534 |
| Q | y | 213 | Asp | GlyArgIleLysProAspValMetAlaProGlyThrTyrIleLeuSerAlaArgSer | 232 |
| D | b | 1535 | GAC | GGCGCCATCAAGCGCGAGCTCACCGCTCCCGCGGTGGACATCATCGCGCTCGGCG | 1594 |
| Q | y | 233 | Ser | LeuAlaProAspSerSerPheThrAlaAsnHisAspSerLysTyrAlaTyrMetGly | 252 |
| D | b | 1595 | GAG | GCNACACATCGCGCAGGAGTCCGTGAGGACCGCGCGCTACATGACCATCTCC | 1654 |
| Q | y | 253 | Gly | ThrSerMetAlaThrProIleValAlaGlyAsnValaGlnLeuArgGluHisPhe | 272 |
| D | b | 1655 | GG | CACGTGATGGCAGCCCGCACGTCCGCGGCGCGCGCTCTCTGAAGCAGCAG--- | 1711 |
| Q | y | 273 | Val | LysAsnArgGlyIleThrProLysProSerLeuLeuLysAlaAlaLeu----- | 289 |
| D | b | 1712 | ----- | CACCCCGACTGGACCTCCGCGCACTGAAGGCGCG | 1747 |
| Q | y | 290 | Ile | AlaGlyAlaAlaAspValGlyLeuGlyTyrProAsnGlyAsnGlnGlyTrpGlyArg | 309 |
| D | b | 1748 | CT | CACCGGCTCCACCAAGGCGGC-----AAGTACACCCCGTTCGAGCAGGTTTCGG | 1804 |
| Q | y | 310 | Val | ThrLeuAspLysSerLeuAsnValAlaTyrValAsnGlnSerSerAlaLeuSer--- | 328 |
| D | b | 1805 | AT | CAGCGCGACAGGGCTCCAGCAGACGTCGATCCGCCGCCCGTCTCGTGAGCTTC | 1864 |
| Q | y | 329 | ----- | -----ThrSerGlnLysAlaIleTyrThr | 336 |
| D | b | 1865 | GG | CTCCAGAGTGGCGCGCACCGCAGCAGCGCGTCCACCAAGCAGCTGACCTACCG | 1924 |
| Q | y | 337 | Phe | ThrAlaThrAlaGlyLysProLeuLysIleSerLeuValTrpSerAsp----- | 353 |
| D | b | 1925 | AAC | TCGCGCACCCAGGACGTCACTGAAGTCAAGTCGACCGCCACCGACCCCAAGGCG | 1984 |
| Q | y | 354 | ----- | AlaProAlaSer-----ThrThrAlaSerValThrLeuValAsnAspLeuAsp | 369 |
| D | b | 1985 | AAG | CGCGCGCGCGGGGTCTTTCACGCTCGGCGCCACCAACG----- | 2026 |
| Q | y | 370 | Leu | ValIleThrAlaProAsnGlyThrArgTyrValGlyAsnAspPheSerAlaProPhe | 389 |
| D | b | 2027 | ----- | GTGACCGTCCCGGGGGCGGC-----ACGCGCTCGGTC | 2059 |
| Q | y | 390 | Asp | AsnAsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAsnSerProGlnSer | 409 |
| D | b | 2060 | GAC | ATGACCGCGCACACCGCG-----CTCGCGCGCAGCGGTGGAC | 2098 |
| Q | y | 410 | Gly | ThrTyrThrIle----- | 414 |
| D | b | 2099 | GG | CGGTACTCGGCGTACGTGGTCGCCACGGCGCGGCGAGCGGTCCGACGGCGCGC | 2158 |
| Q | y | 415 | ----- | GluValGlnAlaTyrAsnValProVal | 423 |
| D | b | 2159 | CG | GTGACGCGGAGGTTCGACTCGTACGACGTGACCGTCC | 2197 |

Patent No. 6361987
GENERAL INFORMATION:
APPLICANT: AKIRA ARISAWA et al.
TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC
HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-DIHYDROPYRIDINE
ITS EXPRESSION PRODUCT
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
STREET: 2033 K Street, N.W., #800
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20006
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/514,340
FILING DATE: 28-Feb-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/000,016
FILING DATE: January 30, 1998
ATTORNEY/AGENT INFORMATION:
NAME: Lee Cheng
REGISTRATION NUMBER: 40,949
REFERENCE/DOCKET NUMBER: <Unknown>
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-721-8200
TELEFAX: 202-721-8250
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2539 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
ORIGINAL SOURCE:
ORGANISM: Streptomyces viridosporus
STRAIN: A-914
FEATURE:
NAME/KEY: CDS
LOCATION: 338...2539
IDENTIFICATION METHOD: E
SEQUENCE DESCRIPTION: SEQ ID NO: 3:
MS-09-514-340-3

[illegible]

QY 62 AlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlyAla 81
DB 1112 GCGGCGGACAAAGTGGCGCACCGCACCGCTCGCTCGATCGCGGGGACCGGGCGCC 1171
QY 82 ThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIle 97
DB 1172 CAGTCCAAAGGCAAGTACAGGCGCTCGCACCGCGCGGATCCTCAACGCGCAAGTGC 1231
QY 96 MetAspSerSerGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGln 117
DB 1232 CTCGACGACTCGGGT-----TTCGGCGACGACTCCGGCATCCTCGCGGATGAGTGG 1285
QY 118 AlaPheSerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAla 137
DB 1286 GCGGCGGCGGCGGCGGCGGCTCGTCCATGACCTGGGC-----GGCATG 1333
QY 138 TyrThrThrAspSerArgAsnValAspAspTyrValArgLys-----AsnAspMetThr 155
DB 1334 GACACACGGAGACCGCGCTGGAGCGGGCTCGACAAGCTCTCGCGCGAGAAGGCG 1393
QY 156 IleLeuPhe-----AlaAlaGlyAsnGlyGlyProAsnGlyGlyThrIleSerAlaPro 173
DB 1394 GTCTGTTCGCCATCGCGCGCGGCGGACGAGGCGCGGAG-----TCGATCGTTCGCC 1447
QY 174 GlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGly 193
DB 1448 GCGAGCGCGGACCGCGGCTCACCGCTCGCGGC-----1480
QY 194 SerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGlyPro---ThrIys 212
DB 1481 -----GTGAGCAAGACAAAGCTCGCGGCTCTCTCCACCGCGCGGCTCGCGC 1534
QY 213 AspGlyArgIleLeuProAspValMetAlaProGlyThrTyrIleLeuSerAlaArgSer 232
DB 1535 GACGGCGCATCAAGCGGACGTCACCGCTCCCGCGGTGGACATCACGCGCGCTCGCGC 1594
QY 233 SerLeuAlaProAspSerSerPheTrpAlaAsnHisAspSerLysTyrAlaTyrMetGly 252
DB 1595 GAGGCAACACATCGCGCAGAGGTCTGTGAGGACCGCGCGGTATCATGACCATCTCC 1654
QY 253 GlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPhe 272
DB 1655 GGCACGCTCATGCGGACCGCGGCGGCGGCGGCGGCGGCTCTCTGAAGCAGCAG--- 1711
QY 273 ValLysAsnArgGlyIleThrProLysProSerLeuLeuLysAlaAlaLeu----- 289
DB 1712 -----CACCGCGACTGGACCTCGCGGAACTGAAGGCGCG 1747
QY 290 IleAlaGlyAlaAlaAspValGlyLeuGlyTyrProAsnGlyAsnGlnGlyTrpGlyArg 309
DB 1748 CTCACCGGCTCCACCAAGGCGGC---AAGTACACCGCGTTCGAGCAGGTTTCGGGCGCG 1804
QY 310 ValThrLeuAspLysSerLeuAsnValAlaTyrValAsnGlnSerSerAlaLeuSer--- 328
DB 1805 ATCCAGGCGGACAAAGCGCTCCAGCAGACCGTGTATCGCGGACCGCGGCTCTCGGTGAGCTTC 1864
QY 329 -----ThrSerGlnLysAlaThrTyrThr 336
DB 1865 GCGGTCCAGAGTGGCGGCGACCGACGAGCGCGGTCACCAAGAGCGTACCTACCGC 1924
QY 337 PheThrAlaThrAlaGlyLysProLeuLysIleSerLeuValTrpSerAsp----- 353
DB 1925 AACCTCGGACCCAGGACGTCAGCTGAAGCTGACGTGACCGCGGCGGCGGCGGCGGCGG 1984
QY 354 -----AlaProAlaSer-----ThrThrAlaSerValThrLeuValAsnAspLeuAsp 369
DB 1985 AAGCGGCGCGCGGCGGCTTCTTCACGCTGGGCGGCGGCGGCGGCGGCGGCGGCGG 2026
QY 370 LeuValIleThrAlaProAsnGlyThrArgTyrValGlyAsnAspPheSerAlaProPhe 389
DB 2027 -----GTGACCGCTCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2059
QY 390 AspAsnAsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAsnSerProGlnSer 409

DB 2060 GACATGACCGCGACACCGG-----CTCGGGGCGGCGGCGGAC 2098
QY 410 GlyThrTyrThrIle-----414
DB 2099 GCGCGCTACTCTCGCGGTACGTGTGTCGCCAGCGGCGGCGGCGGCGGCGGCGGCGG 2158
QY 415 -----GluValGlnAlaTyrAsnValProVal 423
DB 2159 GCGGTGACGCGGAGGTGCGAGTCTGACGAGTACGAGTACCGTC 2197
RESULT 14
US-09-000-016-1
Sequence 1, Application US/09000016
Patent No. 6143541
GENERAL INFORMATION:
APPLICANT: AKIRA ARISAWA et al.
TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC
TITLE OF INVENTION: HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-DIHYDROPYRIDINE DERI
TITLE OF INVENTION: ITS EXPRESSION PRODUCT
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
STREET: 2033 K Street, N.W., #800
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20006
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/000,016
FILING DATE: January 30, 1998
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Warren M. Cheek, Jr.
REGISTRATION NUMBER: 33,367
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-721-8200
TELEFAX: 202-721-8250
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2809 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
ORIGINAL SOURCE:
ORGANISM: Streptomyces viridosporus
STRAIN: A-914
ORIGINAL SOURCE:
ORGANISM: Streptomyces antibioticus
STRAIN:
FEATURE:
NAME/KEY: CDS
LOCATION: 338...2539
IDENTIFICATION METHOD: E
FEATURE:
NAME/KEY: CDS
LOCATION: 2540...2809
IDENTIFICATION METHOD: P
US-09-000-016-1
Alignment Scores: 6.23e-27 Length: 2809
Pred. No.:

| | | | |
|---|--------|--|-----|
| Score: | 354.50 | Matches: | 139 |
| Percent Similarity: | 41.49% | Conservative: | 56 |
| Best Local Similarity: | 29.57% | Mismatches: | 173 |
| Query Match: | 15.79% | Indels: | 103 |
| DB: | 3 | Gaps: | 17 |
| US-09-985-689A-2 (1-434) x US-09-000-016-1 (1-2809) | | | |
| QY | 2 | AspValAlaArgGlyLeuValValAlaAspValAlaGlnSerSerTyGlyLeuTyArgly 21 | |
| DB | 953 | GACAGCTCGTGGCGAGATCGGCGCCGCCAAGCGTGTCCGCC---GGCTACGACGGC 1009 | |
| QY | 22 | GlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 41 | |
| DB | 1010 | AAGGCGTGAAGATCGCGCTCTGGACACCGGTGTCCGACGAGC----- 1054 | |
| QY | 42 | MetHisGluAlaPheArgGlyLysIleThrAlaLeuTyAlaLeuGlyArgThrAsnAsn 61 | |
| DB | 1055 | ---CATCCGACCTCAAGGCGCGGTGACCGCGTCCAAAGAACTTCACCGCGCGCCCGGC 1111 | |
| QY | 62 | AlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlyAla 81 | |
| DB | 1112 | GCCGCGCAAGGTGGCGGCGACGACCGACCGATCGCTCGATCGCGCGGCGACGGCGCC 1171 | |
| QY | 82 | ThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIle 97 | |
| DB | 1172 | CAGTCCAAAGGCAAGTACAGGCGGTGCGACCGCGCGCGGATCCTCAACGGCAAGGTC 1231 | |
| QY | 98 | MetAspSerSerGlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGln 117 | |
| DB | 1232 | CTCGACGACTCCGGT-----TTCCGCGACACCTCCCGCATCTCCGCGCGCATCGAGTGG 1285 | |
| QY | 118 | AlaPheSerAlaGlyAlaArgIle-HisThrAsnSerTyrGlyAlaAlaValAlaGlyAl 137 | |
| DB | 1286 | GCGCGCGCGAGGCGCGCGTCTCAACATGAGCTGGCGCGCATCGACACCGCGAG 1345 | |
| QY | 137 | aTyrThrThrAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLe 157 | |
| DB | 1346 | ACCGCGCGTGGAGCGCGCG-GTCGACAAAGCTGTCCGCGAGAGGCGGTCTGTGTTCG 1404 | |
| QY | 157 | pHeAlaAlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLy 177 | |
| DB | 1405 | CATCGCGCGCGCAACGAGGCGCGGAG-----TCGATCGGTTCGCGCGCGAGCGCGGA 1458 | |
| QY | 177 | sAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyAlaAs 197 | |
| DB | 1459 | GCGCGCGCTCACCGTCCGCGC-----GTCGA 1485 | |
| QY | 197 | pAsnIleAsnHisValAlaGlnPheSerSerArgGlyPro---ThrLysAspGlyArgI 216 | |
| DB | 1486 | CGACGAGCAGAGCTCGCGCGATCTCTCTCCACCGCGCCCGCTCGCGGCGCGCCAT 1545 | |
| QY | 216 | eLysProAspValMetAlaProGlyThrTyIleLeuSerAlaArgSerSerLeuAlaPr 236 | |
| DB | 1546 | CAAGCGCGACGTACCGCTCCCGCGGTGACATCACGCGCGCTCCGCGAGGCGCAACGA 1605 | |
| QY | 236 | oAspSerSerPheThrAlaAsnHisAspSerLysTyAlaTyMetGlyGlyThrSerMe 256 | |
| DB | 1606 | CATCGCGCAGAGGTGCGTGAGGACCGCGCGGTATACATGACCATCTCCGCGACGTGAT 1665 | |
| QY | 256 | tAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnAr 276 | |
| DB | 1666 | GCGCACCGCGACGTGCGCGCGCGCGCGCGCTCTCTGAGCAGCAG----- 1711 | |
| QY | 276 | gGlyIleThrProLysProSerLeuLeuLysAlaAlaLeu-----IleAlaGlyAl 293 | |
| DB | 1712 | -----CACCCCGACTGACCTCCGCGCGAACTGAAGGCGCGGTCTCACCGGCTC 1758 | |
| QY | 293 | aAlaAspValGlyLeuGlyTyProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAs 313 | |
| DB | 1759 | CACCAGGCGCGC---AGTACACCGCTTCGAGCAGGGTTCGGCGCGGATCCAGGCGGA 1815 | |
| QY | 313 | plySerLeuAsnValAlaTyValAsnGluSerSerAlaLeuSer----- 328 | |

RESULT 15

US-09-514-340-1
 ; Sequence 1, Application US/09514340
 ; Patent No. 6361987
 ; GENERAL INFORMATION:
 ; APPLICANT: Akira ARISAWA et al.
 ; TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC
 ; HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-DIHYDROXYRIDINE D
 ; ITS EXPRESSION PRODUCT

NUMBER OF SEQUENCES: 7
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
 STREET: 2033 K Street, N.W., #800
 CITY: Washington
 STATE: D.C.
 COUNTRY: U.S.A.
 ZIP: 20006
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: MS-DOS
 SOFTWARE: Wordperfect 5.1
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/514,340
 FILING DATE: 28-Feb-2000
 CLASSIFICATION: <Unknown>
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 09/000,016
 FILING DATE: January 30, 1998
 ATTORNEY/AGENT INFORMATION:
 NAME: Lee Cheng
 REGISTRATION NUMBER: 40,949
 REFERENCE/DOCKET NUMBER: <Unknown>
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 202-721-8200
 TELEFAX: 202-721-8250
 TELEX: <Unknown>
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2809 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double

```

; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Streptomyces antibioticus
; STRAIN: <Unknown>
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 338...2539
; IDENTIFICATION METHOD: E
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2540...2809
; IDENTIFICATION METHOD: P
; SEQUENCE DESCRIPTION: SEQ ID No: 1:
US-09-514-340-1

Alignment Scores:
Pred. No.: 6,23e-27 Length: 2809
Score: 354.50 Matches: 139
Percent Similarity: 41.4% Conservative: 56
Best Local Similarity: 29.57% Mismatches: 173
Query Match: 15.79% Indels: 103
DB: 4 Gaps: 17

US-09-985-689A-2 (1-434) x US-09-514-340-1 (1-2809)
QY 2 AspValAlaArgGlyLeuValValAlaAspValAlaGlnSerSerTyrGlyLeuTyrGly 21
DB 953 GACACGTCTCGCGGAGATCGCGCCGCCCAAGGGGTGGTCGGCC---GGCTACGACGGC 1009
QY 22 GlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSer 41
DB 1010 AAGGGGTGAAGATCGCGCTCTGGACACCGGTGTGACACAGC-----1054
QY 42 MetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsn 61
DB 1055 ---CATCCGACCTGAAGGGCGGGGTGACCGGCTCAAGAACTTCCACCGCGCGCGCGC 1111
QY 62 AlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlyAla 81
DB 1112 GCGGGCGCAAGGTGGCCACGGCACCCACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1171
QY 82 ThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIle 97
DB 1172 CAGTCCCAAGGGCAAGTACAAAGGGCGTGGACCGCGCGCGCGCGATCTCAACGCGCAAGGTC 1231
QY 98 MetAspSerSerGlyLeuGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGln 117
DB 1232 CTCGACGACTCCGGT-----TTCGGCGACGACTCCGGCATCTCGCGGCGATGAGTGG 1285
QY 118 AlaPheSerAlaGlyAlaArgIle-HisThrAsnSerTrpGlyAlaAlaValAsnGlyAl 137
DB 1286 GCGGGCGCGGCGGCGCGCGCGTCAACATGAGCTGGCGGCGATGGACACACCGGAG 1345
QY 137 aTyrThrThrAspSerArgAsnValAspTyrValArgLysAsnAspMetThrIleIle 157
DB 1346 ACCGACCCGCTGAGGCGCGG-GTCACAGCTGTCCGCGGAGAGGGCGTCTGTTCGC 1404
QY 157 uPheAlaAlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLy 177
DB 1405 CATCGCGCGCGCAACGAGGCGCGGAG-----TCGATCGTTCGCCCGCGGCGCGGA 1458
QY 177 sAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAs 197
DB 1459 CGCGCGCGCTCACCCTCGCGGCC-----GTGGA 1485
QY 197 pAsnIleAsnHisValAlaGlnPheSerSerArgGlyPro---ThrLysAspGlyArgIle 216
DB 1486 CGACAGGCAAGCTCGCGGACTTCTCTCCACCGGCGCGCGCGCTCGCGGCGCGCCAT 1545
QY 216 eLysProAspValMetAlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAlaPr 236
DB 1546 CAAGCGCGGAGTCAACCGCTCCCGCGGTGGACATCACGGCGCGCTCGCGGCGGCGCAACGA 1605

```

```

QY 236 oAspSerSerPheTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMe 256
DB 1606 CATCGGCCAGGAGGTCTGGTGGAGCGCGCGCTACATGACCATCTCCGGCACGTGAT 1665
QY 256 tAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnAr 276
DB 1666 GCGACCCCGCGACGTGCGCGGCGCGCGCTCTCTGAAGCAGCAG-----1711
QY 276 ggIlyIleThrProLysProSerLeuLeuLysAlaAlaLeu-----IleAlaGlyAl 293
DB 1712 -----CACCCCGACTGGACCTCCGCCGAACCTGAAGGGCGCGCTCACCGGCTC 1758
QY 293 aAlaAspValGlyLeuGlyTyrProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAs 313
DB 1759 CACCAAGGGCGCGC---AAGTACACCCCGGTGAGCAGGGTTCGGCGCGGATCCAGGCCGA 1815
QY 313 pLysSerLeuAsnValAlaTyrValAsnGlnSerSerAlaLeuSer-----328
DB 1816 CAAGGGCTCCAGCAGACCGTGTATCGCCGACCGGCTCTCGGTGAGCTTCGGGCTCCAGCA 1875
QY 329 -----ThrSerGlnLysAlaThrTyrThrPheThrAlaTh 340
DB 1876 GTGGCGCGCACACCGACGAGCGCGTCCACCAAGCAGCTGACCTACCGCAACCTCGGCAC 1935
QY 340 xAlaGlyLysProLeuLysIleSerLeuValTrpSerAsp-----AlaPr 355
DB 1936 CAAGGAGCTCAGCTGAGCTGAGCTGACGTGACCGCCACCGACCCCAAGGGCAAGCGGCCCC 1995
QY 355 oAlaSer-----ThrThrAlaSerValThrLeuValAsnAspLeuAspLeuValIleTh 373
DB 1996 GCGGGGCTTCTTCACGCTGGCGCGCCACCG-----GTGAC 2031
QY 373 xAlaProAsnGlyThrArgTyrValGlyAsnAspPheSerAlaProPheAspAsnAsnTr 393
DB 2032 GTCCCGCGCGGCGCGC-----AGCGCTCGCTCGACATGACCGC 2070
QY 393 pAspGlyArgAsnAsnValGluAsnValPheIleAsnSerProGlnSerGlyThrTyrTh 413
DB 2071 CGACACCGCG-----CTCGGCGGCGACGCTGGACGCGCGGTACTC 2109
QY 413 rIle-----414
DB 2110 GGCGTACGTGTCGCCACGCGCGCGGCGGCGAGCGGTCCGACGCGCGCGCGGTGCGCGG 2169
QY 415 -GluValGlnAlaTyrAsnValProVal 423
DB 2170 CGAGTCTGAGTCTGACGACGTGACCGCTC 2197

```

Search completed: April 4, 2004, 12:01:36
Job time : 92.5977 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 4, 2004, 08:10:21 ; Search time 328.004 Seconds
(without alignment)
4948.852 Million cell updates/sec

Title: US-09-985-689A-2

Perfect score: 2245

Sequence: 1 NDVARGIVKADVAQSSVGLY.....EVQAYNVVPGQNFSLAIVN 434

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5

Ygapop 10.0, Ygapext 0.5

Fgapop 6.0, Fgapext 7.0

Delop 6.0, Delext 7.0

Searched: 2466186 seqs, 1870095128 residues

Total number of hits satisfying chosen parameters: 4932372

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-O=/cgn2_1/USPTO.spool/US09985689/runat_31032004_161809_4271/app.query.fasta_1.3498
-DB=Published Applications NA -OPM=fastap -SUFFFX=trnpb -MINMATCH=0.1
-LOOPL=0 -LOOPEXT=0 UNITS=bits -SPAR=1 -END=1 -MATRIX=blotsum62
-TRANS=human40.cdi -LIST=45 -DOCLALIGN=200 -THR SCORE=pct -THR MAX=100
-THR_MIN=0 -ALIGN=15 -MOD=LOCAL -OUTFMT=ptc -NORM=ext -HAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US09985689 @CGN_1_1.601 @runat_31032004_161809_4271
-NCPU=6 -ICPU=3 -NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGJOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:

1: /cgn2_6/ptodata/1/pubnpa/US07_PUBCOMB.seq:
2: /cgn2_6/ptodata/1/pubnpa/PCT_NEW_PUB.seq:
3: /cgn2_6/ptodata/1/pubnpa/US06_NEW_PUB.seq:
4: /cgn2_6/ptodata/1/pubnpa/US06_PUBCOMB.seq:
5: /cgn2_6/ptodata/1/pubnpa/US07_NEW_PUB.seq:
6: /cgn2_6/ptodata/1/pubnpa/PCTUS_PUBCOMB.seq:
7: /cgn2_6/ptodata/1/pubnpa/US08_NEW_PUB.seq:
8: /cgn2_6/ptodata/1/pubnpa/US08_PUBCOMB.seq:
9: /cgn2_6/ptodata/1/pubnpa/US09A_PUBCOMB.seq:
10: /cgn2_6/ptodata/1/pubnpa/US09B_PUBCOMB.seq:
11: /cgn2_6/ptodata/1/pubnpa/US09C_PUBCOMB.seq:
12: /cgn2_6/ptodata/1/pubnpa/US09_NEW_PUB.seq:
13: /cgn2_6/ptodata/1/pubnpa/US10A_PUBCOMB.seq:
14: /cgn2_6/ptodata/1/pubnpa/US10B_PUBCOMB.seq:
15: /cgn2_6/ptodata/1/pubnpa/US10C_PUBCOMB.seq:
16: /cgn2_6/ptodata/1/pubnpa/US10_NEW_PUB.seq:
17: /cgn2_6/ptodata/1/pubnpa/US60_NEW_PUB.seq:
18: /cgn2_6/ptodata/1/pubnpa/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------|
| ----- | | | | | |

| | | | | | | |
|----|-------|------|---------|----|----------------------|--------------------|
| 1 | 2191 | 97.6 | 1305 | 15 | US-10-385-662-1 | Sequence 1, Appli |
| 2 | 457.5 | 20.4 | 1977 | 13 | US-10-090-624-11 | Sequence 11, Appli |
| 3 | 425.5 | 19.0 | 1236 | 13 | US-10-090-624-2 | Sequence 2, Appli |
| 4 | 425.5 | 19.0 | 1962 | 13 | US-10-090-624-15 | Sequence 15, Appli |
| 5 | 360.5 | 16.1 | 3624 | 14 | US-10-156-761-5701 | Sequence 5701, Ap |
| 6 | 360.5 | 16.1 | 9025608 | 14 | US-10-156-761-1 | Sequence 1, Appli |
| 7 | 345.5 | 15.4 | 3417 | 14 | US-10-156-761-3306 | Sequence 3306, Ap |
| 8 | 345.5 | 15.4 | 9025608 | 14 | US-10-156-761-1 | Sequence 1, Appli |
| 9 | 338.5 | 15.3 | 135838 | 14 | US-10-314-657-1 | Sequence 1, Appli |
| 10 | 308.5 | 15.1 | 1329 | 9 | US-09-974-300-1934 | Sequence 1934, Ap |
| 11 | 308.5 | 13.7 | 1560 | 15 | US-10-084-846A-113 | Sequence 113, App |
| 12 | 308.5 | 13.7 | 59816 | 15 | US-10-084-846A-1 | Sequence 1, Appli |
| 13 | 308.5 | 13.7 | 59816 | 15 | US-10-084-846A-2 | Sequence 2, Appli |
| 14 | 307 | 13.7 | 4765 | 15 | US-10-090-624-5 | Sequence 5, Appli |
| 15 | 283 | 12.6 | 3788 | 10 | US-09-927-827-33 | Sequence 29, Appli |
| 16 | 277 | 12.3 | 3743 | 10 | US-09-927-827-39 | Sequence 39, Appli |
| 17 | 275 | 12.2 | 2192 | 12 | US-10-424-599-112429 | Sequence 112429, |
| 18 | 273 | 12.2 | 1306 | 9 | US-09-966-921A-1 | Sequence 1, Appli |
| 19 | 273 | 12.2 | 1330 | 9 | US-09-966-921A-5 | Sequence 5, Appli |
| 20 | 270 | 12.0 | 3303 | 14 | US-10-156-761-5384 | Sequence 5384, Ap |
| 21 | 261 | 11.6 | 1971 | 9 | US-09-974-300-1935 | Sequence 1935, Ap |
| 22 | 256 | 11.4 | 840 | 14 | US-10-209-812-1 | Sequence 1, Appli |
| 23 | 254 | 11.3 | 2166 | 12 | US-10-344-231-17 | Sequence 17, Appli |
| 24 | 253 | 11.3 | 2166 | 12 | US-10-363-332A-17 | Sequence 17, Appli |
| 25 | 246 | 11.0 | 1485 | 9 | US-09-974-300-1938 | Sequence 1938, Ap |
| 26 | 246 | 11.0 | 1140 | 8 | US-08-322-678-11 | Sequence 11, Appli |
| 27 | 246 | 11.0 | 1140 | 8 | US-08-322-678-12 | Sequence 12, Appli |
| 28 | 246 | 11.0 | 1140 | 16 | US-10-323-324-11 | Sequence 11, Appli |
| 29 | 246 | 11.0 | 1140 | 16 | US-10-323-324-12 | Sequence 12, Appli |
| 30 | 246 | 11.0 | 1143 | 14 | US-10-313-853-6 | Sequence 6, Appli |
| 31 | 246 | 11.0 | 2588 | 12 | US-10-344-231-20 | Sequence 20, Appli |
| 32 | 246 | 11.0 | 2588 | 12 | US-10-363-332A-20 | Sequence 20, Appli |
| 33 | 240.5 | 10.7 | 3884 | 10 | US-09-927-827-34 | Sequence 34, Appli |
| 34 | 236.5 | 10.5 | 1497 | 8 | US-08-322-678-6 | Sequence 6, Appli |
| 35 | 236.5 | 10.5 | 1497 | 9 | US-09-060-854B-1 | Sequence 1, Appli |
| 36 | 236.5 | 10.5 | 1497 | 14 | US-10-033-325-1 | Sequence 1, Appli |
| 37 | 236.5 | 10.5 | 1497 | 15 | US-10-228-572-1 | Sequence 1, Appli |
| 38 | 236.5 | 10.5 | 1497 | 15 | US-10-423-649-1 | Sequence 1, Appli |
| 39 | 236.5 | 10.5 | 1497 | 16 | US-10-323-324-6 | Sequence 6, Appli |
| 40 | 234.5 | 10.4 | 1140 | 15 | US-10-146-905A-9 | Sequence 9, Appli |
| 41 | 234.5 | 10.4 | 1149 | 15 | US-10-146-905A-7 | Sequence 7, Appli |
| 42 | 234.5 | 10.4 | 1494 | 14 | US-10-104-693-1 | Sequence 1, Appli |
| 43 | 233 | 10.4 | 2267 | 12 | US-10-344-231-18 | Sequence 18, Appli |
| 44 | 233 | 10.4 | 2267 | 12 | US-10-363-332A-18 | Sequence 18, Appli |
| 45 | 232.5 | 10.4 | 4338 | 9 | US-09-891-711-3 | Sequence 3, Appli |

ALIGNMENTS

RESULT 1
US-10-385-662-1
; Sequence 1, Application US/10385662
; Publication No. US20040002432A1
; GENERAL INFORMATION:
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SAITO, TSUYOSHI
; APPLICANT: SAITO, KAZUHIRO
; APPLICANT: SUMITOMO, NOBUYUKI
; APPLICANT: IZAWA, YOSHIFUMI
; APPLICANT: SAKI, KAISUHIKA
; APPLICANT: KOBAYASHI, TOHRO
; APPLICANT: NOMURA, MASAFUMI
; TITLE OF INVENTION: Alkaline protease
; FILE REFERENCES: 234938USO
; CURRENT APPLICATION NUMBER: US/10/385,662
; CURRENT FILING DATE: 2003-03-12
; PRIOR APPLICATION NUMBER: JP 2002-081428
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: JP 2002-165987
; PRIOR FILING DATE: 2002-06-06
; PRIOR APPLICATION NUMBER: JP 2002-304230
; PRIOR FILING DATE: 2002-10-18
; PRIOR APPLICATION NUMBER: JP 2002-304231

```
; PRIOR FILING DATE: 2002-10-18
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1305
; TYPE: DNA
; ORGANISM: Bacillus sp. KSM-KP43
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1305)
; OTHER INFORMATION:
US-10-385-662-1

Alignment Scores:
Pred. No.: 1,34e-235 Length: 1305
Score: 2191.00 Matches: 419
Percent Similarity: 99.54% Conservativity: 13
Best Local Similarity: 96.54% Mismatches: 2
Query Match: 97.59% Indels: 0
DB: 15 Gaps: 0

US-09-985-689A-2 (1-434) x US-10-385-662-1 (1-1305)

QY 1 AsnAspValAlaArgGlyLeuValAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
DB 1 AATGATGTTGGCGTGGATTTGCAAGCGGATGGCTCAGACGAGCTACGGGTTGTAT 60
QY 21 GlyGlnGlyGlnLeuValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
DB 61 GGAACAAGGACAGATCGATGCGGTGCGGATACAGGCGCTTGATACAGGTGCGCAATGACAGT 120
QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
DB 121 TCGATCGATGAAGCTTCGCGGGAATAATCTGCAATATATGCAATGGGACGAGAT 180
QY 61 AsnAlaAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
DB 181 AATGCCAATGATACGAATGCTGATGTACGCATGTGGCTGGCTCGCTATTAGAAACGCG 240
QY 81 AlaThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
DB 241 TCCACTANTAAAGAAATGGCGCTCAGCGGAATCTAGTCTTCAATCTATCATGTATGAC 300
QY 101 SerGlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaPheSer 120
DB 301 GGTGGGGGACTTGGAGGACTTACCTCGAATCTGCAAACTTATGCAAGCCTTATGAGCCAGCATA 360
QY 121 AlaGlyAlaArgIleHisThrAsnSerTyrGlyAlaAlaValAsnGlyAlaTyrThr 140
DB 361 GCTGGTGCCAGAAATTCATCAAACTCTGGGGACGACAGTGAATGGGCTTACACACA 420
QY 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
DB 421 GATTCAGAAATGTGGATGACTATGTGCGCAAAATGATATGACGATCTTTTCGCTGCC 480
QY 161 GlyAsnGlyGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAlaIle 180
DB 481 GGGAAATGAAGCCGACCGCGAAATCATAGTGCACACGACGACGACGACGACGACGACGAC 540
QY 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
DB 541 ACAGTCGAGTACGGAAACCTCCGCCCAAGCTTTGGGTCTTATCGGACAAATATCAAC 600
QY 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
DB 601 CATGTGCGCAGTTCTCTTCAGTGGACCGACAAAGATGGAGGATGCAACCGGATGTC 660
QY 221 MetAlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerPhe 240
DB 661 ATGCAACGGGAAGCTTCACTATACGCAAGATCTCTCTTCGACCGGATCTCTCTCTTC 720
QY 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260
```

```
DB 721 TGGCGAACCATGACAGTAAATATGATACATGCGGTGGAACTCCATGCGTACACGGATC 780
QY 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
DB 781 GTTGTCTGGAACGTCGACAGCTTCGTGAGCATTTTGTGAAACACAGAGGCATCACACCA 840
QY 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspValGlyLeuGlyTyr 300
DB 841 AAGCCTTCTCTATTAAAGCGGACCTGATGCGGGTGCAGCTGACATCGGCTTGGCTAC 900
QY 301 ProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
DB 901 CCGAACGGTAACCAAGATGGGACGAGTGACATTGATAAATCCCTGAACGTTGCTAT 960
QY 321 ValAsnGluSerSerAlaLeuSerThrSerGlnLysAlaThrTyrThrPheThrAlaThr 340
DB 961 GTGAACGAGTCCAGTTCTCTATCCACAGCCAAAAGACGACTCTGTTTACTGCTACT 1020
QY 341 AlaGlyLysProLeuLysIleSerLeuValTyrSerAspAlaProAlaSerThrAla 360
DB 1021 GCGGCAAGCCTTTGAAATCTCCCTGGTATGCTGATGCCCCCTGCGAGCACACTGCT 1080
QY 361 SerValThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrArgTyr 380
DB 1081 TCCGTAAACGCTTGTCAATGATCTGGACCTTGTCATTACCCCTCCAAATGCGACACAGTAT 1140
QY 381 ValGlyAsnAspPheSerAlaProPheAspAsnAsnTyrAspGlyArgAsnValGlu 400
DB 1141 GTAGGAATGACTTACTTCTGCGCATACATGATACTGGATGGCGGCAATACGTAGAA 1200
QY 401 AsnValPheIleAsnSerProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
DB 1201 AATGTATTATTAAATGACCAACAGCGGACGCTATACAAATTGAGGTACAGGCTTATAAC 1260
QY 421 ValProValGlyProGlnAsnPheSerLeuAlaIleValAsn 434
DB 1261 GTACCGTTGGACACAGACTTCTGTTGGCAATTGTGAT 1302

RESULT 2
US-10-090-624-11
; Sequence 11, Application US/10090624
; Publication No. US20020132335A1
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA-6
; CURRENT APPLICATION NUMBER: US/10/090,624
; CURRENT FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: 09/445,472
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 11
; LENGTH: 1977
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-10-090-624-11

Alignment Scores:
Pred. No.: 6.75e-41 Length: 1977
Score: 457.50 Matches: 136
Percent Similarity: 44.66% Conservativity: 69
Best Local Similarity: 29.63% Mismatches: 153
Query Match: 20.38% Indels: 101
DB: 13 Gaps: 17
```


US-09-985-689A-2 (1-434) x US-10-090-624-11 (1-1977)

QY 8 ValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyrGlyGlnGlyGlnLeuValAla 27
DB 433 ATAGGGCCGATACCGTCTGAACCTCTCGCTACGACGAGCGGTGTGGTGGCC 492
QY 28 ValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGluAlaPheArg 47
DB 493 ATCGTCGATACGGGTATAGACGGAAC-----CACCCCGATCTGAAG 534
QY 48 GlyLysIleThrAlaLeuTyr-----AlaLeuGlyArgThrAsnAsnAlaAsnAspThrAsn 66
DB 535 GCAAGGTCATAGCGTGTAGACCGCTCAACCGCAGGTCCGACCGCCCTACGATGACCAAG 594
QY 67 GlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlyAlaThrAsnLys----- 84
DB 595 GGACACGGAAACCCACGTTGGGGTATCGTTCGGGAACCGGACGGGTAACTCCAGTAC 654
QY 85 ---GlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerSer 101
DB 655 ATAGCGTCCGCGCGCGGCGAAGTCGTCGGCGTCAAGGTTCTCGGTGCGGACGGTTCG 714
QY 102 GlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaPheSerAla 121
DB 715 GGAAGCGTCTCCACCATCATCGCGGTGTGTACTGGTCTGTCAGAACAGGACCAAGTAC 774
QY 122 GlyAlaArgIle-----GlyHisThrAsnSer 129
DB 775 GGGATAAGGGTCAACCTCTCCCTCGCTCTCCAGAGTCCGACGGAACCGCACTCC 834
QY 130 TrpGlyAlaAlaValAsnGlyAlaTyrThrThrAspSerArgAsnValAspTyrVal 149
DB 835 CTCAGTCAGCGCTCAACACCCCTGGGACGCC----- 867
QY 150 ArgLysAsnAspMetThrIleLeuPheAlaAlaGlyAsnGluGlyProAsnGlyGlyThr 169
DB 868 -----GGTATAGTAGTCTCGTCGCGCGGCAACAGCGGCGGCAACACCTACACC 918
QY 170 IleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArg 189
DB 919 GTCCGCTACCCCGCGCGGAGGAGGTCTAAGCGTCAATACCGTCGGTGCA----- 963
QY 190 ProSerPheGlySerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGly 209
DB 964 -----GTTACAGCAACGCAACATCGCAGCTTCTCCAGCAGGGGA 1005
QY 210 ProThrLysAspGlyArgIleLysProAspValMetAlaProGlyThrTyrIleLeuSer 229
DB 1006 CCGACCGCGGACGGAAGGCTCAAGCGGAAAGTGTGCGCCCGCGGCTTGACATCATAGCC 1065
QY 230 AlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsnHisAspSerLysTyrAla 249
DB 1066 CCGCGCGCCAGC-----CGAACCCAGTGGGCGCCCGGATTAACGACTACTACACC 1116
QY 250 TyrMetGlyGlyThrSerMetAlaThrProIleValAlaGly-----AsnValAlaGlnLeu 268
DB 1117 AAGGGCTCTGGAACCAAGCATGCGCCACCGCGTTCGGGCGTTGGCGGCTCATCCTC 1176
QY 269 ArgGluHisPheValLysAsnArgGlyIleThrProLys-----ProSerLeuLeuLys 286
DB 1177 CAGGCGCCAC-----CCGAGCTGAGCCCGGCAAGGTGAAG 1212
QY 287 AlaAlaLeuIleAlaGlyAla-----AlaAspValGlyLeu 298
DB 1213 ACCGCGCTCATCGAGACCGCCGACATAGTCGCGCCCAAGGAGATAGCGGACATCGCCTAC 1272
QY 299 GlyTyrProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeu----- 316
DB 1273 GTGGCG-----GGTAGGGTGAACGCTCAAGGCGCCCTCAAGTAC 1311
QY 317 ---AsnValAlaTyrValAsnGluSerSerAlaLeuSerThrSerGlnLysAlaThrTyr 335
DB 335

1312 GACGACTACGCCAAGCTCACCTTCCGCTCCGCGCACAGGAGGCCACCCAC 1371
QY 336 ThrPheThrAlaThrAlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaPro 355
DB 1372 ACCTTCGAGCTCAGCGGCCACCTTCGAGCGCCACCTCTACTGGGAC----- 1422
QY 356 AlaSerThrAlaSerValThrLeuValAsnAspLeuAspLeuValIleThrAlaPro 375
DB 1423 -----ACGGCTCGAGCGACATCGACCTCTACCTACGACCC 1461
QY 376 AsnGlyThrArgTyrValGlyAsnAspPheSerAlaProPheAspAsnAsnTrpAspGly 395
DB 1462 ACGGGAACGAG-----GTTGACTACTCTTACCGGCC 1494
QY 396 ArgAsnAsnValGluAsnValPheIleAsnSerProGlnSerGlyThrTyrThrIleGlu 415
DB 1495 TACTACGGCTTCGAGAAGGTGCGCTACTACAACCGCGGACCGCGGACCTGGAGGTCAAG 1554
QY 416 ValGluAlaTyrAsnValProValGlyProGlnAsnPheSerLeuAlaIleValAsn 434
DB 1555 GTCGTGAGCTACAG-----GGCGGCGGAACCTACCGGTGACGCTGTCAGC 1602

RESULT 3
US-10-090-624-2
; Sequence 2, Application US/10090624
; Publication No. US20020132335A1
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikuneshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCES: TAKAKURA-6
; CURRENT APPLICATION NUMBER: US/10/090,624
; CURRENT FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: 09/445,472
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 1236
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-10-090-624-2

Alignment Scores:
Pred. No.: 1,3e-37 Length: 1236
Score: 425.50 Matches: 140
Percent Similarity: 42.77% Conservative: 61
Best Local Similarity: 29.79% Mismatches: 147
Query Match: 18.95% Indels: 122
DB: 13 Gaps: 19

US-09-985-689A-2 (1-434) x US-10-090-624-2 (1-1236)

QY 12 ValAlaGlnSerSerTyrGlyLeuTyr-----GlyGlnGlyGln 24
DB 24 GTCTGCGACTCAAGTTATGCAACTTACGTTTGGAACTGGGATATGATGGTCTGGAAT 83
QY 24 nIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGln 44
DB 84 CACAATAGGAATAATTTGACACTGGAATTGAC-----GCTTCTCATCC 125
QY 44 uAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAsn-- 63
DB 126 AGATCTCCAAAGGAAAGTA-----ATTGGGTGGGTAGATTGTCATGG 170
QY 64 -----AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGln 78


```
Db 747 GTGGCGCTGTGATACAAAGATAAGTACGGAATTAAGTCTAATTAATCTTCTCTGTGTC 806
Qy 132 a
Db 807 AAGCCAGAGCTCAGATGCTACTGACGCTCTAAGTCAGGCTGTAAATGACGGTGGATGC 866
Qy 140 rAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrLeuPheAla 160
Db 867 T-----CGATTAGTGTGTGTGTGTGCGC 890
Qy 160 aGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAla 180
Db 891 TGGAAACAGTGGACCTTCAAGTACAAATCGGTTCTCCAGCAGCTGCAACCAAGTAT 950
Qy 180 eThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAs 200
Db 951 TACAGTTGGAGCC-----GTTGACCAAGTATGA 977
Qy 200 nHisValAlaGlnPheSerArgGlyProThrLysAspGlyArgIleLysProAspVa 220
Db 978 TGTATATAACAGCTTCTCAGCAGAGGCGCAACTGCAGACGCGAGCTTAAGCTGAGGT 1037
Qy 220 lMetAlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPh 240
Db 1038 TGTTCCTCCAGGAACCTGGATAAATGCTGCCAGAGCAAGT-----GGAACCTAGCAT 1088
Qy 240 eTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIl 260
Db 1089 GGTCAACCAATTATGACTATTACACAGCAGCTCTGGGACATCAATGGCAATCCTCA 1148
Qy 260 eValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPr 280
Db 1149 CGTAGCTGTATTGACGCGCTCTTCTGCTCAA-----GCACACCC 1187
Qy 280 oLys-----ProSerLeuLeuLysAlaAlaLeuIleAlaGlyAla----- 293
Db 1189 GAGCTGAGCTCAGCAAAAGCAAAAGCCCTCATAGAACTGCTGATATCGTAAGCC 1247
Qy 294 -----AlaAspValGlyLeuGlyTyrProAsnGlyAsnGlnGlyTyrGlyArgVa 310
Db 1248 AGATGAAATAGCCGATATAGCTACGCTGCA-----GGTAGGGT 1286
Qy 310 lThrLeuAspLysSerLeuAsnValAlaTyrValAsnGluSerSerAlaLeuSerThrSe 330
Db 1287 TAATGCATACAAAGCTATAAAC-----TACGATAACTATGCAAGAGTAGTTCACATGG 1340
Qy 330 rGlnLysAla-----ThrTyrThrPheThrAlaThrAlaGlyLysProLe 345
Db 1341 ATATGTTGCCAAAGGCGAGCCAACTCACCAGTTCTGTTATTAGCGGAGCTTCTGTCGT 1400
Qy 345 uLysIleSerLeuValTrpSerAspAlaProIleSerThrThrAlaSerValThrLeuVa 365
Db 1401 AACTGCCACATTATATCTGGACAATGCCAAT----- 1431
Qy 365 lAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrArgTyrValGlyAsnAspPh 385
Db 1432 -AGCGACCTTGATCTTACCTCTACGATCCCATGGAAACCAAG----- 1473
Qy 385 eSerAlaProPheAspAsnAsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAs 405
Db 1474 -----GTTGACTACTCTTACACCGCTACTATGATTCGAAAGGTTGTTATTA 1523
Qy 405 nSerProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsnValProValGlyPr 425
Db 1524 CAACCCAACTGATGGAACTGACAAATTAAGTTGTAGCTACAGC-----GGAAG 1574
Qy 425 oGlnAsnPheSerLeuAlaIleValAsn 434
Db 1575 TGCAAACTATCAAGTAGATGTGTGAAGT 1602
```

RESULT 5

US-10-156-761-5701

```
; Sequence 5701, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 5701
; LENGTH: 3624
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(3624)
US-10-156-761-5701
```

| | | | |
|------------------------|----------|---------------|------|
| Alignment Scores: | 1,26e-29 | Length: | 3624 |
| Pred. No.: | 360.50 | Matches: | 129 |
| Score: | 44.39% | Conservative: | 53 |
| Percent Similarity: | 31.46% | Mismatches: | 161 |
| Best Local Similarity: | 16.06% | Indels: | 67 |
| Query Match: | 14 | Gaps: | 17 |
| DB: | | | |

US-09-985-689a-2 (1-434) x US-10-156-761-5701 (1-3624)

```
Qy 8 ValLysAlaAspValAlaGlnSerSer-----Tyr 17
Db 556 GTCGAGGCGCACATGTCGCGGAGCAACGCGCAGATCGGTACGCGGCGCGCGTGGGAGCC 615
Qy 18 GlyLeuTyrGlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArg 37
Db 616 GGCTCAGCGGCGACGCGCTCACCCTGCGCGTGCTGCACACGCGGCTGCACACC----- 669
Qy 38 AsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGly 57
Db 670 -----ACTCACCCCGACCTCGCGGCGCGGCTGTCGCGGAGCAAGAGCTTCATC 717
Qy 58 ArgThrAsnAsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeu 77
Db 718 GACGGGAGGAGTCCGCCACCGCAACGCGCACCGGACCCACGTCACTCGACCTCGCGC 777
Qy 78 GlyAsnGlyAla-----ThrAsnLysGlyMetAlaProGlnAlaAsnLeuVal 93
Db 778 GGCAGCGGCGCGCTCCGACGCGCAGGCGCGGCGTCCGCGCGGTGCCACGCTCGCC 837
Qy 94 PheGlnSerIleMetAspSerSerGlyGlyLeuGlyLeuProSerAsnLeuGlnThr 113
Db 838 GTCGGCAAGTGTCTCAGCAGCACCGACCGCGG-----GGAAGCAGATCCCGATC 885
Qy 114 LeuPheSerGlnAlaPheSerAla-----GlyAlaArgIleHisThrAsnSerTrp 130
Db 886 ATCGCGGCGATGGAATGGCGCGCGCGGAGCTGCGGCCAGGATCGTCTCGATGAGCCTC 945
Qy 131 GlyAla-----AlaValAsnGlyAlaTyrThrThrAsp-----SerArgAsnValAsp 146
Db 946 GGATCGACCGAGCGCAGCGG-----ACCGACCCCATGCCCGCGCGCTCGAC 996
Qy 147 AspTyrValArgLysAsnAspMetThrIleLeuPheAlaAlaGlyAsnGluGlyProAsn 166
Db 997 ACCCTCTCCGAGGAGACCGCGCGCTCTCTGTCGCGCGGAGCAACACCGCTGCC--- 1053
```

```
167 GlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAlaThrGlu 186
1054 CCCTCTCGATCGGTCGGCCGCGCGGAGTCTCGGCGTACCGTCCGGCGCC----- 1107
187 AsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsnHisValAlaGlnPheSer 206
1108 -----GTCGACTCATCCGACCGCGCGCTACTTCCAC 1140
207 SerArgGlyProThrLys---AspGlyArgIleLysProAspValMetAlaProGlyThr 225
1141 ACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1200
226 TyrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsnHisAsp 245
1201 GACATCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1245
246 SerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIleValAlaGlyAsnVal 265
1246 -----TACCTCCATGACGGTACGTGATGGCGAGCGCCCATGTGCGGGGGTCCGC 1299
266 AlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrProLysProSerLeuLeu 285
1300 GCGCTCTCCGCGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1359
286 LysAlaAla-----LeuIleAlaGlyAlaAlaAspValGlyLeuGlyTyrProAsnGly 303
1360 TCACGTCCGAGCAACTCGACCGCTCCGTATATCATGTTGGGGCGGGTCCGGTCAGTGT 1419
304 AsnGlnGlyTrpGly---ArgValThrLeuAspLysSerLeuAsnValAlaTyrValAsn 322
1420 CCGAGCGCGTCCGCGCGCGCGTCCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1479
323 GluSerSerAlaLeuSerThrSerGlnLysAlaThrTyrThrPheThrAlaThrAlaGly 342
1480 TGCGCCCATGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1539
343 LysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrThrAlaSerVal 362
1540 ACGACGGTCGAGTTCGAGCTCGCGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1590
363 ThrLeuValAsnAspLeuAspLeuValIleThrAlaPro-----AsnGlyThr 378
1591 ACCCTCGCC-----GACACCGCACTCACCGTCCGCGCGCGCGCGCGCGCGCGCG 1641
379 ArgTyrValGlyAsnAspPheSerAlaPro 388
1642 ACCGTACCGCGCGCGCGCTCCAGGCTCCG 1671

RESULT 6
US-10-156-761-1
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
```

| | | | |
|--|---------|--|---------|
| QY | 266 | AlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrProLysProSerLeuLeu | 285 |
| Db | 6919557 | GCCTCTCTCCCGAGCAGCACCCCGACTGGACCGCGCGCGGTCAAGGACGCGCTGATG | 6919616 |
| QY | 286 | LysAlaAla-----LeuIleAlaGlyAlaAlaAspValGlyLeuGlyTyrProAsnGly | 303 |
| Db | 6919617 | TCCACGTCCGAGCAACTCGAGCGCTCCGTATATCAGTTGGGGCGGGTCGGGTCACTGTG | 6919676 |
| QY | 304 | AsnGlnGlyTrpGly---ArgValThrLeuAspLysSerLeuAsnValAlaTyrValAsn | 322 |
| Db | 6919677 | CCGAGACCGCTCGCGCGCGCTCACCGCACCGGCGAGCCGACCTCGGCTTCCACCGC | 6919736 |
| QY | 323 | GluSerSerAlaLeuSerThrSerGlnLysAlaThrTyrThrPheThrAlaThrAlaGly | 342 |
| Db | 6919737 | TGGCCCCATAGCGCCGATCGACCGCTCACAAGACGGTCACTACTCCAACTCTCTCGAC | 6919796 |
| QY | 343 | LysProLeuLysLeuValTrpSerAspAlaProAlaSerThrThrAlaSerVal | 362 |
| Db | 6919797 | ACGACGGTCGAGTTGAGCTTCGCGTGGCGGCGCGCGCGGTGC-----GCC | 6919847 |
| QY | 363 | ThrLeuValAsnAspLeuAspLeuValIleThrAlaPro-----AsnGlyThr | 378 |
| Db | 6919848 | ACCTTCGCC-----GACACCGCACTCACCGTGCCTCCGCCACCGCACCGCGCCACC | 6919898 |
| QY | 379 | ArgTyrValGlyAsnAspPheSerAlaPro | 388 |
| Db | 6919899 | ACCGTAGCGCGCAGCGCTCCAAGGCTCCG | 6919928 |
| RESULT 7 | | | |
| US-10-156-761-3306 | | | |
| ; Sequence 3306, Application US/10156761 | | | |
| ; Publication No. US20030119018A1 | | | |
| ; GENERAL INFORMATION: | | | |
| ; APPLICANT: OMURA, SATOSHI | | | |
| ; APPLICANT: IKEDA, HARUO | | | |
| ; APPLICANT: ISHIKAWA, JUN | | | |
| ; APPLICANT: HORIKAWA, HIROSHI | | | |
| ; APPLICANT: SHIBA, TADAYOSHI | | | |
| ; APPLICANT: SAKAKI, YOSHIYUKI | | | |
| ; APPLICANT: HATTORI, MASAHIRA | | | |
| ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES | | | |
| ; FILE REFERENCE: 249-262 | | | |
| ; CURRENT APPLICATION NUMBER: US/10/156,761 | | | |
| ; CURRENT FILING DATE: 2002-05-29 | | | |
| ; PRIOR APPLICATION NUMBER: JP 2001-204089 | | | |
| ; PRIOR FILING DATE: 2001-05-30 | | | |
| ; PRIOR APPLICATION NUMBER: JP 2001-272697 | | | |
| ; PRIOR FILING DATE: 2001-08-02 | | | |
| ; NUMBER OF SEQ ID NOS: 15109 | | | |
| ; SEQ ID NO 3306 | | | |
| ; LENGTH: 3417 | | | |
| ; TYPE: DNA | | | |
| ; ORGANISM: Streptomyces avermitilis | | | |
| ; FEATURE: | | | |
| ; NAME/KEY: CDS | | | |
| ; LOCATION: (1)..(3417) | | | |
| US-10-156-761-3306 | | | |
| Alignment Scores: | | | |
| Pred. No.: 5,57e-28 Length: 3417 | | | |
| Score: 345.50 Matches: 130 | | | |
| Percent Similarity: 43.68% Conservative: 53 | | | |
| Best Local Similarity: 31.03% Mismatches: 158 | | | |
| Query Match: 15.39% Indels: 80 | | | |
| DB: 14 Gaps: 13 | | | |
| US-09-985-689A-2 (1-434) x US-10-156-761-3306 (1-3417) | | | |
| QY | 18 | GlyLeuTyrGlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArg | 37 |
| Db | 745 | GGGTACAGCGCAGCGCGTCAAGATCGCGTCTCTGACACCGGTGTCAC----- | 795 |
| QY | 38 | AsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGly | 57 |

| | | | |
|----|------|--|------|
| Db | 796 | -----GCGACCCACCGGACCTCAGGACGAGTGGCGGAGTCCAGAACTTCTCC | 846 |
| Qy | 58 | ArgThrAsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeu | 77 |
| Db | 847 | GCGCGCGGACGCCCGACCACTTCGGTCA CGGCACGACGCTCGCGTCCATCGCGCG | 906 |
| Qy | 78 | GlyAsnGlyAlaThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuVal | 93 |
| Db | 907 | GGACCGCGGCCCAAGTCCACGCAAGTACAAAGGTGTGCGCGCGGCGGACGATCTCT | 966 |
| Qy | 94 | PheGlnSerIleMetAspSerSerGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThr | 113 |
| Db | 967 | AACGGCAAGTCTCGAGACACGGC-----TCCGGCGACGACTCCGCGATCTTGGCC | 1020 |
| Qy | 114 | LeuPheSerGlnAlaPheSerAlaGlyAlaArgIle-HisThrAsnSerTrpGlyAlaAl | 133 |
| Db | 1021 | GGCATGGAGTGGCGCGCGAGCGCGCGACGCTCGTCAACCTGAGCTTGGCGCGCGGC | 1080 |
| Qy | 133 | avalAsnGlyAlaTyThrThrAspSerArgAsnValAspAspTyrValArgLysAsnAs | 153 |
| Db | 1081 | GACACCCCGAGATCGACCGCTGGAGCGGAG-GTCAACAAGCTCTCCGAGGAGAGGG | 1139 |
| Qy | 153 | pMetThrIleLeuPheAlaAlaGlyAsnGluGlyProAsnGly---GlyThrIleSerAl | 172 |
| Db | 1140 | CATCTCTTCGCGATCGCCCGCGCACGAAGCGAGTTCGGGAGCAGACCATCGGCTTC | 1139 |
| Qy | 172 | aproglyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPh | 192 |
| Db | 1200 | CCGCGGAGCGCGCGCACCGCTCACCGTCCGCGCGC----- | 1236 |
| Qy | 192 | eglySerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGlyProThrLy | 212 |
| Db | 1237 | -----GTGAACGACGACGACAGCTGGCGTCTCTCCAGCGCGCGCCCGGCT | 1286 |
| Qy | 212 | sAspGlyArgIleIleysProAspValMetAlaProGlyThrTyrIleLeuSerAlaArgSe | 232 |
| Db | 1287 | GGACGGCGCCATCAAGCCGACGTCAACCGCACCGCGGTGGAGATCACC CGCGCC | 1341 |
| Qy | 232 | rSerLeuAlaProAspSerSerPheTrpAlaAsnHisAspSerLys-----TyrAl | 249 |
| Db | 1342 | -----GCGGCGCGGGCAGCGTCATCGACCAGGAGGTCGGCGAGAGCGGACGCTACCT | 1397 |
| Qy | 249 | aTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuAr | 269 |
| Db | 1398 | CACCATCTCGGTACGTGATGGCGACCCCGCATGTGCGCGGCGCGCCGCGATCTCTCAA | 1457 |
| Qy | 269 | gGluHisPheValIleAsnArgGlyIleThrProLysProSerLeu-----LeuLysAl | 287 |
| Db | 1458 | CGACGACAC-----CCCACTGGTCTGTCGCCAGGCTCAAGGG | 1496 |
| Qy | 287 | aAlaLeuIleAlaGlyAlaAlaAspValGlyLeuGlyTyrProAsnGlyAsnGlnGlyTr | 307 |
| Db | 1497 | CGGCTACCGGCTCCGGAAG-----GGCGGCAAGTACACGCGGTTCCAGCAGGCGCTC | 1550 |
| Qy | 307 | pGlyArgValThrLeuAspLysSerLeuAsnValAlaTyrValAsnGluSerSerAlaLe | 327 |
| Db | 1551 | GGGCGGATCGCGTTCGACAGGCGATCAAGAGTCCGTGATCGCCAACCCGGAACCTCGGT | 1610 |
| Qy | 327 | userThrSerGlnLysAlaThrTyrThrPhe-ThrAlaThrAlaGlyLysProLeuLysI | 347 |
| Db | 1611 | GAGC-----TTCCGATCCAGTGGCGGCACACCGACGA | 1646 |
| Qy | 347 | leserLeuValTrpSerAspAlaProAlaSerThrThrAlaSerValThrLeuValAsnA | 367 |
| Db | 1647 | CAAGCCGGTCACCCAGGAGCTACCTACCGCACCTCGGACGAGTGACGTC-ACGCTGA | 1705 |
| Qy | 367 | spLeuAspLeuValIleThrAlaProAsnGlyThrArgTyrValGlyAsnAspPheSerA | 387 |
| Db | 1706 | ACCTCGGCTCGACGGCACCAACCCCAAGGGCGTC-----GCCG | 1744 |
| Qy | 387 | laProPheAspAsnAsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAsnSerP | 407 |

RESULT 9
US-10-314-657-1
; Sequence 1, Application US/10314657
; Publication No. US20030175888A1
; GENERAL INFORMATION:
; APPLICANT: SHEN, Ben
; APPLICANT: CHENG, Yi-Qiang
; APPLICANT: TANG, Gong-Li
; TITLE OF INVENTION: Discrete Acyltransferases Associated with Type I Polyketide
; TITLE OF INVENTION: Synthases and Methods of Use
; FILE REFERENCE: 054030-0021
; CURRENT APPLICATION NUMBER: US/10/314,657
; CURRENT FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: PCT/US02/08937
; PRIOR FILING DATE: 2002-03-22

```
; PRIOR APPLICATION NUMBER: US 60/278,935
; PRIOR FILING DATE: 2001-03-26
; NUMBER OF SEQ ID NOS: 214
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 1
; LENGTH: 135638
; TYPE: DNA
; ORGANISM: Streptomyces atroolivaceus
US-10-314-657-1

Alignment Scores:
Pred. No.: 2,58e-25 Length: 135638
Score: 343.00 Matches: 128
Percent Similarity: 43.06% Conservative: 58
Best Local Similarity: 29.63% Mismatches: 144
Query Match: 15.28% Indels: 102
DB: 14 Gaps: 17

US-09-985-689A-2 (1-434) x US-10-314-657-1 (1-135638)
Qy 6 GlyLeuValLysAlaAspValAlaGlnSer----- 15
Db 8067 GGCAGGGTGAAGCGCATCTCGCGACTCCACCGCCGAGATCGCGCGGCGAGAGGTATGG 8126

Qy 16 SerTyrGlyLeuTyrGlyGlnGlyGlnLeuValAlaValAlaAAspThrGlyLeuAspThr 35
Db 8127 GCGAGGGCGACACCGCGGAGAGGTGAGGTGCGGATGCTCGACAGCGCGCGGACACC 8186

Qy 36 GlyArgAsnAspSerSerMetHisGluAlaPheArgGlyLysile-----ThrAlaLeu 53
Db 8187 -----GAACACCGGACCTGTCGGCGAGGTGTCGACAGCGCGCAGC 8228

Qy 54 TyrAlaLeuGlyArgThrAsnAsnAlaAsnAspThrAsnGlyHisGlyThrHisValAla 73
Db 8229 TTCGTCCCGCGGAGGACGATCGCC---GACTACACGGCCACCGGACGCGATCGCC 8285

Qy 74 GlySerValLeuGlyAsnGlyAlaThrAsn-----LysGlyMetAlaProGln 89
Db 8286 TCGACCATGTCGGCAGCGGACGCCCTCCGACGCGAAGGAGCGGGTGTGCGCTCCGCG 8345

Qy 90 AlaAsnLeuValPheGlnSerIleMetAspSerSerGlyGly----- 103
Db 8346 GCCCGCGTGTCCGTCCGGAAGGTGCTCAACTCCGAGGCGACGCCAGGATCGTGGATC 8405

Qy 104 LeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaPheSerAlaGlyAla 123
Db 8406 ATCGCGGCGATGGAG-----TGGCGCGCGCGGACCAAGGCC 8444

Qy 124 ArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyrThrAsp----- 141
Db 8445 AGGATCATCAGCATGAGCTGGGC-----GCGCGCGGTGACAAGACGACCCGATG 8495

Qy 142 SerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAlaGly 161
Db 8496 AGCCAGGCGGTGACAACTAGCCACGACGCGCGGTGTCGATGCGCGCGGCG 8555

Qy 162 AsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThr 181
Db 8556 AACGCGCGCGGCAC-----TCCATCAGACCGCGGTGCGGACGACTCGCGCTGACC 8609

Qy 182 ValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsnHis 201
Db 8610 GTCGCGCGCC-----GTCGACTCCACCGGACG 8636

Qy 202 ValAlaGlnPheSerSerArgGlyProThrLys---AspGlyArgIleLysProAspVal 220
Db 8637 CTCGCGGACTTCTCCAGCAGGCGCGGTGACGCGGCGGCGGTGAAGCGGAGATC 8696

Qy 221 MetAlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
Db 8697 ACCGCGCGCGGCGTGCATATCGTCGCGCGGCGTGCATCAAGCGCGGCTCCGCTAC 8756

Qy 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260
Db 8757 -----TACACCATGATGCGGCGACGTCGATGCGGACGCGGCAC 8795

Qy 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
Db 8796 GTCGCGGCGGTGCGCGCGCTCTCCGCGCGAGACCCCGACTGCGGCGGCGACCGCTC 8855

Qy 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspValGlyLeuGlyTyr 300
Db 8856 AAGGAGGCGACTGTCAGCAGCGCCCAAGCAAGCGCGCG-----TAC 8897

Qy 301 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
Db 8898 ACCCGGTACCGCGCGCGCGCGCGGTGCGACGCGCGCGCGCGGTGCGACACCGCTC 8957

Qy 321 ValAsnGlnSerSerAlaLeuSer-----ThrSer 330
Db 8958 TTCGCCACACGACCGCGCTACTCCGCTTCCACACGTGGCCCCCGGAGCGGAGACC 9017

Qy 331 GlnLysAlaThrTyrThrPheThrAlaThrAlaGlyLysProLeuLysIleSerLeuVal 350
Db 9018 GATGTCGCGACGCTGACGTACACCAACGTCGCGCGACGCCCGGTGACGCTCAACCTG 9077

Qy 351 TrpSer-----AspAlaPro 355
Db 9078 GTCAAGGCGACCGTCCCGCGCGGTGTGTACGCTCTCCGAGGACCATGTACCGTGCC 9137

Qy 356 AlaSerThrAlaSerValThrIleuValAsnAspLeuAsp----- 369
Db 9138 GCGCAGCGCACCGCCACGCTGACCGCGGTCTGGAACAAGTGGCGGCGGCGACGAG 9197

Qy 370 -----LeuValIleThrAlaProAsnGlyThr 378
Db 9198 TCGGTGAGCGCGGTGATCATCCGTCAGGACGCGACG 9233

RESULT 10
US-09-974-300-1934
; Sequence 1934, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; TITLE OF INVENTION: Expression
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; PRIOR FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1934
; LENGTH: 1329
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-1934

Alignment Scores:
Pred. No.: 8.3e-28 Length: 1329
Score: 338.50 Matches: 105
Percent Similarity: 43.53% Conservative: 43
Best Local Similarity: 30.88% Mismatches: 109
Query Match: 15.08% Indels: 83
DB: 9 Gaps: 13

US-09-985-689A-2 (1-434) x US-09-974-300-1934 (1-1329)
Qy 11 AspValAlaGlnSerSerTyrGlyLeuTyrGlyGlnIleValAlaAlaAsp 30
Db 406 GAATGCTGAGAAACAATCAGCGTGCAGCGCAAGAGGAGTGCAGATCGCTGTCTTGTAT 465
```

31 ThrGlyLeuAspThrGlyArgAsnAspSerMetHisGluAlaPheArgGlyLysile 50
466 AGCGGGTA-----TACCCTCAGAGATCTTGAAGGACGAGGATC 504
51 ThrAlaLeuTyr---AlaLeuGlyArgThrAsnAlaAsnAspThrAsnGlyHisGly 69
505 AGGGCTTTTCAAGACTTTTCAACCCAGAGAACAGAACCTATGATGACAAATGGGCACGCG 564
70 ThrHisValAlaGlySerValLeuGlyAsnGlyAlaAlaThrAsn-----LysGly 85
565 ACACACTGCGCGGTGATCTTGGGAACCGGCGGCTCATCGGTCAGTACCGCGGA 624
86 MetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSerSerGlyGlyLeuGly 105
625 CCGTCTCTCTGAAGCAGACTTGTGCGTGTAAAGTATTGGACAAAATGGGATCC---GGA 681
106 GlyLeuProSerAsnLeuGln-----SerTyrGlyAlaAlaValAsnGly 136
682 TCGCTGAAACCGTCAATTAAGCGTAGATTGGTGCAATTCATTCAATAAGGAATCCT 741
113 -----ThrLeuPheSerGlnAlaPheSerAlaGlyAlaArgIleHisThrAsn-128
742 GATGATCCGATCGACATTATTCAATGTCATTGGGTGCAGAACCTTGGCTACGAGAAAT 801
129 -----SerTyrGlyAlaAlaValAsnGly 136
802 GAAGAAAGATCCAGTCGTGTAAGCTGTTATGCGAGCATGGACCGCGGCAAT----- 855
137 AlaTyrThrThrAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIle 156
856 -----GTTGTA 861
157 LeuPheAlaAlaGlyAsnGlyProAsnGlyGlyThrIleSerAlaProGlyThrAla 176
862 TGTGCGGACCGCGCACTCGGCTCTGTATGCGCAACCAATTCGCGCGGCGGTGTACG 921
177 LysAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAla 196
922 AGCAAGATTATTACAGTCGAGCGCTTGGATGAC---AGGATACACTCAGCGCGGAGGAT 978
197 AspAsnIleAsnHisValAlaGlnPheSerArgGlyProThrLysAspGlyArgIle 216
979 GACGAT-----GTCGCTCTTATTCAAGCAGAGGCGGCAATCTATGTCGAAGTC 1029
217 LysProAspValMetAlaProGlyThrTyrIleLeuSerAlaArgSer-----SerLeu 234
1030 AACCGGACTGTGTTGACCGGACAAATATTACGTCGCTGTTCCCGGATCTTT 1089
235 AlaProAspSerSerPheTyrAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThr 254
1090 CTCGATAAGCTGCAAAAACAAACAGAGTCGCGCACAATAATATGATGTCGCGAACC 1149
255 SerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLys 274
1150 TCGATGCTACGCGATCTGCGGAGGAATTCGCGCA-----CTTATCCTT 1194
275 AsnArgGlyIleThrProLysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAla 294
1195 CAGCAAGCCCGGCGACAGAACTGTGTAAGTCAAAACAGCTGTCTAATGGACGCT---ACC 1251
295 AspVal-----GlyLeuGlyTyrProAsnGlyAsnGln 305
1252 GATTATTGGAAGATCGCGATCCAAATGTTTACGTCGAGGGTACATCAACGACGACAA 1311

RESULT 11
US-10-084-846A-113
; Sequence 113, Application US/10084846A
; Publication No. US2004006026A1
; GENERAL INFORMATION:
; APPLICANT: WEITNAUER, GABRIELE
; APPLICANT: MUELENWEG, AGNES
; APPLICANT: TREFFER, AXEL
; APPLICANT: BECHTHOLD, ANDREAS

TITLE OF INVENTION: AVILAMYCIN DERIVATIVES
FILE REFERENCE: 1974-005
CURRENT APPLICATION NUMBER: US/10/084,846A
CURRENT FILING DATE: 2003-02-25
PRIOR APPLICATION NUMBER: PCT/EP01/09815
PRIOR FILING DATE: 2001-08-24
PRIOR APPLICATION NUMBER: DE 101 09 166.4
PRIOR FILING DATE: 2001-02-25
NUMBER OF SEQ ID NOS: 120
SOFTWARE: PatentIn Ver. 3.2
SEQ ID NO 113
LENGTH: 1560
TYPE: DNA
ORGANISM: Streptomyces viridochromogenes
FEATURE:
OTHER INFORMATION: avix16 dna: partial sequence of coding strand 1; nucleotide 1
OTHER INFORMATION: corresponds to nucleotide 56,198 of coding strand 1.
US-10-084-846A-113

Alignment Scores:
Pred. No.: 2,44e-24 Length: 1560
Score: 308.50 Matches: 115
Percent Similarity: 39.34% Conservative: 51
Best Local Similarity: 27.25% Mismatches: 159
Query Match: 13.74% Indels: 97
DB: 15 Gaps: 17

US-09-985-689A-2 (1-434) x US-10-084-846A-113 (1-1560)

QY 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
DB 451 GGGCAGGAGTGCAGCGGCTACGTATCGACACCGCGGCTC-----CGC 492
QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyr---AlaLeuGlyArgThr 59
DB 493 ATCACCACACGACACTTCGCGCGCGGCTCTCTACGGCTACGACGCCATCGACCAACGAC 552
60 AsnAsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn 79
553 AACACCGCCAGGACGCGCCACGCGCAGCGCAGCGCTGCGCGGACGCGTCCGCGGCAAC 612
QY 80 GlyAlaThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAsp 99
613 GCC-----TACGCGTCCGCAAGAGCCAGATCTAGGCTGCGCGTCTGCTGAC 663
QY 100 SerSerGly-----GlyLeuGlyGlyLeuProSerAsn 110
664 AACTCGCGCCAGGACACCCAGCGCGGCTGTCGCGGCTGCGCTGCGCGGCAAC 723
QY 111 LeuGlnThrLeuPheSerGlnAlaPheSerAlaGlyAlaArgIleHisThrAsnSerTyr 130
724 -----GCCGTCAGCGCGCGCTGCGCAACATGTCCTC 756
QY 131 GlyAlaAlaValAsnGlyAlaTyrThrThrAspSerArgAsnValAspAspTyrValArg 150
757 GCGCGCGCGCGACAGCGGCTCGACACGCGGCTGACCAAC-----GCCATG 804
QY 151 LysAsnAspMetThrIleLeuPheAlaAlaGlyAsnGlyProAsnGlyGlyThrIle 170
805 GCCTCCGCGCTCCTTCGCGCGCGGCGGACGAGTCGACCAACGCTCCACGAGG 864
QY 171 SerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAlaThr-----GluAsnLeu 188
865 TCA---CCGCGACGCGTCCAGGAGCCATCACGTCGCGGCGGACGACGCTCGGACGCC 921
QY 189 ArgProSerPheGlySerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArg 208
922 AAGCGCGGCTACTCCAACTACGCTCCGCTCCTC----- 954
QY 209 GlyProThrLysAspGlyArgIleLysProAspValMetAlaProGlyThrIleLeu 228
955 -----GACCTCTTCGCGCGCGGCTGCTCATCATCACC 984

| | | | |
|------|----|---|------|
| 229 | QY | SerAlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsnHisAspSerIysTyr | 248 |
| | | | |
| 985 | Db | TCGGCC | |
| | | | |
| | | ---TGGAACTCAAGCAGACTCGGGAC | 1014 |
| 249 | QY | AlaTyrMetGlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeu | 268 |
| | | | |
| | | ---TGGAACTCAAGCAGACTCGGGAC | |
| 1015 | Db | AACACCATCTCCGGTAGCTGATGGCAGCCCCCAGCTGGCGGGCGCGCGGCTC | 1071 |
| 269 | QY | ArgGluHisPheValIysAsnArgGlyIleThrProLys | 285 |
| | | | |
| | | ---CACCTCGCGCAACCCCTCGGCACCCCGTCCAGGTGCGCAGCGGCGCTGACG | 1125 |
| 1072 | Db | -----CACCTCGCGCAACCCCTCGGCACCCCGTCCAGGTGCGCAGCGGCGCTGACG | 1125 |
| 286 | QY | LysAlaAlaLeuIleAlaGlyAlaAlaAspValGlyLeuGlyTyrProAsnGlyAsnGln | 305 |
| | | | |
| | | ---TCCGCGCGCCACCAACCGCGCTGTCACCAACCCCGCAGCGGCTCGGCCAAC | 1176 |
| 1126 | Db | -----TCCGCGCGCCACCAACCGCGCTGTCACCAACCCCGCAGCGGCTCGGCCAAC | 1176 |
| 306 | QY | GlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyrValAsnGlu | 323 |
| | | | |
| | | ---CGGCTCTGTAGCTCGCGCGCGGCAC | 1203 |
| 1177 | Db | -----CGGCTCTGTAGCTCGCGCGCGGCAC | 1203 |
| 324 | QY | -----SerSerAlaLeuSerThrSerGlnLysAlaThrTyrThrPheThrAla | 339 |
| | | | |
| | | ---GACCACTTCGGGCGCGCGCTTCGAGAACACACCGGTGACTACAGATCAGCGCAACTCC | 1263 |
| 1204 | Db | -----GACCACTTCGGGCGCGCGCTTCGAGAACACACCGGTGACTACAGATCAGCGCAACTCC | 1263 |
| 340 | QY | ThrAlaGlyLysProLeuLysIleSerLeuValTyrSerAspAlaProAlaSerThrThr | 359 |
| | | | |
| | | ---ACGGTCGAGTCCCGGTGACGGTCTCGGCGTCTCCGGCAACGGCCCTCGGCCCTCGCC | 1323 |
| 1264 | Db | -----ACGGTCGAGTCCCGGTGACGGTCTCGGCGTCTCCGGCAACGGCCCTCGGCCCTCGCC | 1323 |
| 360 | QY | AlaSerVal | |
| | | | |
| | | ---ThrLeuValAsnAspLeuAspLeuValIleThrAlaPro | 375 |
| 1324 | Db | -----ThrLeuValAsnAspLeuAspLeuValIleThrAlaPro | 375 |
| 376 | QY | AsnGlyThrArgTyrValGlyAsnAspPheSerAlaProPheAspAsnAsnTyrAspGly | 395 |
| | | | |
| | | ---GAGCGCAGCGGTACACGCTCAAGTCTACGGCAC | 1437 |
| 1384 | Db | -----GAGCGCAGCGGTACACGCTCAAGTCTACGGCAC | 1437 |
| 396 | QY | ArgAsnAsnValGluAsnValPheIleAsnSer | 414 |
| | | | |
| | | ---ProGlnSerGlyThrThrIle | 414 |
| 1438 | Db | -----ProGlnSerGlyThrThrIle | 414 |
| 415 | QY | GluVal | 416 |
| | | | |
| | | ---ATCAACACCAAGTACTCGGTGAACCGCTCTCTCGGAGCGCGGCACCGCAGTGGAACTG | 1497 |
| 1498 | Db | -----ATCAACACCAAGTACTCGGTGAACCGCTCTCTCGGAGCGCGGCACCGCAGTGGAACTG | 1497 |

RESULT 12

US-10-084-846A-1

03-10-001-040A-1
; Sequence 1, Application US/10084846A

; sequence 1; Application No. US20040006026A1

; EDUCATION NO: 0020040000000000
: GENERAL INFORMATION:

GENERAL INFORMATION:
APPLICANT: WEITNAUER, GABRIELE

APPLICANT: WEI INAOER, GABRIELE
APPLICANT: MUHLLENWEG, AGNES

APPLICANT: MUEHLENWEG, AGNES
APPLICANT: TREEZER, AXEL

APPLICANT: IREZER, RAEL
APPLICANT: BECHTHOLD, ANDREAS

; APPLICANT: BECHTOLD, ANDREAS
; TITLE OF INVENTION: AVILAMYCIN

TYPE OF INVENTION: AVIATION DERIVATIVES
FILE REFERENCE: 1974-005

FILE REFERENCE: 1574-003
CURRENT APPLICATION NUMBER: US/10/084.846

; CURRENT APPLICATION NUMBER: US/10/084,849
 : CURRENT FILING DATE: 2003-02-25

: CURRENT FILING DATE: 2003-02-
 : PRIOR APPLICATION NUMBER: PCT/

; PRIOR APPLICATION NUMBER: PCT/EP01/09815
 : PRIOR FILING DATE: 2001-08-24

: PRIOR FILING DATE: 2001-08-24
 : PRIOR APPLICATION NUMBER: DE 1

; PRIOR APPLICATION NUMBER: DE 1
 ; PRIOR FILING DATE: 2001-02-25

;; PRIOR FILING DATE: 2001-02-25
: NUMBER OF SEQ ID NOS: 130

```

; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: Patent In Ver 3.3

```

```

; SOFTWARE: PatentIn Ver. 3.2
; SEQ ID NO 1

```

SEQ ID NO 1
ENCUTY. 50016

LENGTH: 59816
TYPE: DNA

TYPE: DNA
ORGANISM: Streptomyces violaceodromedusae

US 10 084 845 A 1
; ORGANISM: Streptomyces viridochromogenes

US-10-084-846A-1

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

Alignment Scores:

| | | | |
|---|--------|--|-------|
| Best Local Similarity: | 27.25% | Mismatches: | 159 |
| Query Match: | 13.74% | Indels: | 97 |
| DB: | 15 | Gaps: | 17 |
| US-09-985-689A-2 (1-434) x US-10-084-846A-1 (1-59816) | | | |
| Qy | 21 | GlyClnGlyClnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer | 40 |
| Db | 56648 | GGGACGGAGTACGGCGTACGACACCGCGCTC-----CGC | 56689 |
| Qy | 41 | SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyr---AlaLeuGlyArgThr | 59 |
| Db | 56690 | ATCACCCACAGGACTTCGGCGCGCGCTCTACGGCTACGACGCATCGACACGAC | 56749 |
| Qy | 60 | AsnAsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn | 79 |
| Db | 56750 | AAACCGCGCCAGACGGCCACGGCCACGACGACGACGTCGGCGGCGACGGTCGCGGCAAC | 56809 |
| Qy | 80 | GlyAlaThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAsp | 99 |
| Db | 56810 | GCC-----TACGGCGTCGCCAAGAGGCCAAGATCGTAGGGCTCGCGTCTGAAC | 56860 |
| Qy | 100 | SerSerGly-----GlyLeuGlyGlyLeuProSerAsn | 110 |
| Db | 56861 | AACTCCGGCCAGGGCACACCGCCAGGTCGTGCGCGCATCGACTGGTGGTCGCGGGAAC | 56920 |
| Qy | 111 | LeuGlnThrLeuPheSerGlnAlaPheSerAlaGlyAlaArgIleHisThrAsnSerTrp | 130 |
| Db | 56921 | -----GCCGTCAAGCGCGCGTCCGCCAACATGTCCTCCCTC | 56953 |
| Qy | 131 | GlyAlaAlaValAsnGlyAlaTyrThrThrAspSerArgAsnValAspTyrValArg | 150 |
| Db | 56954 | GGCGGGCGGCCGACACGCGCCCTCGACACGGCGGTACGCAAC-----GCCATG | 57001 |
| Qy | 151 | LysAsnAspMetThrIleLeuPheAlaAlaGlyAsnGlyProAsnGlyGlyThrIle | 170 |
| Db | 57002 | GCCTCCGGCGTCACTTCGCCGCGCGCACAGTCCACCGCTCCACGAGG | 57061 |
| Qy | 171 | SerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAlaThr-----GluAsnLeu | 188 |
| Db | 57062 | TCA---CCGCGCGCGTACCGGAGCCATCACGGTCGGCGGACGACGCTCGGACGCC | 57118 |
| Qy | 189 | ArgProSerPheGlySerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArg | 208 |
| Db | 57119 | AAGCGCGGTACTCCAACTACGCTCCGTCCTC----- | 57151 |
| Qy | 209 | GlyProThrLysAspGlyArgIleLysProAspValMetAlaProGlyThrTyrIleLeu | 228 |
| Db | 57152 | -----GACCTCTCGCCCCCGGTCTGTCATCACC | 57181 |
| Qy | 229 | SerAlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsnHisAspSerLysTyr | 248 |
| Db | 57182 | TCGGCC-----TGGAATCAAGCGAUCTCGGCGGACC | 57211 |
| Qy | 249 | AlaTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeu | 268 |
| Db | 57212 | AACACCATCCCGGTACGTGATGGGACCCCGACGTCGTGGCGGCGCGCGCGCTC--- | 57268 |
| Qy | 269 | ArgGluHisPheValLysAsnArgGlyIleThrProLys-----ProSerLeuLeu | 285 |
| Db | 57269 | -----CACCTCGCCGCAACCCCTCGGCCACCCCGTCCAGGTCCGACGCGCTGACG | 57322 |
| Qy | 286 | LysAlaAlaLeuIleAlaGlyAlaAlaAspValGlyLeuGlyTyrProAsnGlyAsnGln | 305 |
| Db | 57323 | TCGCGCGCCACACCGCGGTCTGTCACCAACCCCGGACGGGTCTGCCCAAC----- | 57373 |
| Qy | 306 | GlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyrValAsnGlu----- | 323 |
| Db | 57374 | -----CGGCTCCTGTACGTCCGCGCGCGGCGAC | 57400 |
| Qy | 324 | -----SerSerAlaLeuSerThrSerGlnLysAlaThrTyrThrPheThrAla | 339 |
| Db | 57401 | GACCACCTCCGGGCGCGGTTCAGAAACACGGGTGACTACACATCAGCGACACTCC | 57460 |

QY 340 ThrAlaGlyLysProLeuLysSerLeuValTrpSerAspAlaProAlaSerThrThr 359
 Db 57461 ACGGTGAGTCCCGGTGAGGTCTCCGGCGTCTCCGCAACGGCCCTCGGCCCTCGCC 57520
 QY 360 AlaSerVal-----ThrLeuValAsnAspLeuAspLeuValIleThrAlaPro 375
 Db 57521 GTAGAGTCCATCGTCCACAGTACATCGCGACCTCCAGTCCAGTCCAGTCCAGTCC 57580
 QY 376 AsnGlyThrArgTyrValGlyAsnAspPheSerAlaProPheAsnAsnTrpAspGly 395
 Db 57581 GACGGCAGCGGTACAGCTCAAGTGTACGGCACC-----GGCGGCGAGTTCGGACAAC 57634
 QY 396 ArgAsnAsnValGluValPheIleAsnSer---ProGlnSerGlyThrThrThrIle 414
 Db 57635 ATCAACACACAGTACTCGGTGAGCGCTCTCGAGGGCGGCCAACGGCAAGTGGAAATG 57694
 QY 415 GluVal 416
 Db 57695 CGGGTG 57700
 RESULT 13
 US-10-084-846A-2/c
 ; Sequence 2, Application US/10084846A
 ; Publication No. US2004006026A1
 ; GENERAL INFORMATION:
 ; APPLICANT: WEITNAUER, GABRIELE
 ; APPLICANT: MUHLNUEG, AGNES
 ; APPLICANT: TREFFER, AXEL
 ; APPLICANT: BECHTHOLD, ANDREAS
 ; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES
 ; FILE REFERENCE: 1974-005
 ; CURRENT APPLICATION NUMBER: US/10/084,846A
 ; CURRENT FILING DATE: 2003-02-25
 ; PRIOR APPLICATION NUMBER: PCT/EP01/09815
 ; PRIOR FILING DATE: 2001-08-24
 ; PRIOR APPLICATION NUMBER: DE 101 09 166.4
 ; PRIOR FILING DATE: 2001-02-25
 ; NUMBER OF SEQ ID NOS: 120
 ; SOFTWARE: Patent in Ver. 3.2
 ; SEQ ID NO 2
 ; LENGTH: 59816
 ; TYPE: DNA
 ; ORGANISM: Streptomyces viridochromogenes
 US-10-084-846A-2
 Alignment Scores:
 Pred. No.: 5,64e-22 Length: 59816
 Score: 308.50 Matches: 115
 Percent Similarity: 39.34% Conservative: 51
 Best Local Similarity: 27.25% Mismatches: 159
 Query Match: 13.74% Indels: 97
 DB: 15 Gaps: 17
 US-09-985-689A-2 (1-434) x US-10-084-846A-2 (1-59816)
 QY 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
 Db 3169 GGGCAGGAGTACGGCGGTACGTATCGACACCGCGTC-----CGC 3128
 QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyr---AlaLeuGlyArgThr 59
 Db 3127 ATCACCACAGCGATTCCGGCGCGCGCTCTTACGGCTACGCGCATCGCAACAGCAC 3068
 QY 60 AsnAsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn 79
 Db 3067 AACACCGCCAGGAGCGCCACGGCAGCGACGCGTGGCGCGCGGTTCGGCGCAAC 3008
 QY 80 GlyAlaThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAsp 99
 Db 3007 GCC-----TACGGCGTCCCAAGAACGCAAGATCGTAGGCGTCCGCGTGTGAAC 2957
 QY 100 SerSerGly-----GlyLeuGlyGlyLeuProSerAsn 110

Db 2956 AACTCCGGCAGGCGACACCGCCAGGTCGTCCGCGCATCGACTGGGTCCCGCGAAC 2897
 QY 111 LeuGlnThrLeuPheSerGlnAlaPheSerAlaGlyAlaArgIleHisThrAsnSerTrp 130
 Db 2896 -----GCCGTCAAGCGCGCGTCCGCCAACATGTCCCTC 2864
 QY 131 GlyAlaAlaValAsnGlyAlaTyrThrAspSerArgAsnValAspAspTyrValArg 150
 Db 2863 GGGGGCGCGCGACACCGCCCTCGACACGGCGGTACGCAAC-----GCCATG 2816
 QY 151 LysAsnAspMetThrIleLeuPheAlaAlaGlyAsnGluGlyProAsnGlyGlyThrIle 170
 Db 2815 GCCTCCGGCTCACTCCCTCCGTCGGCGCGCGCACACGAGTCGACCAACGCCCTCCAGAG 2756
 QY 171 SerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAlaThr-----GluAsnLeu 188
 Db 2755 TCA-----CCGACCGCGTCACCGAGGCGCATCGGTCCGCGCGACCGACGAGTCGCGCGC 2699
 QY 189 ArgProSerPheGlySerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArg 208
 Db 2698 AAGCGCGCTACTCCAACTACGCTCCGCTCCTC----- 2666
 QY 209 GlyProThrLysAspGlyArgIleLysProAspValMetAlaProGlyThrTyrIleLeu 228
 Db 2665 -----GACTCTTCGCCCGCGGTTCGTCCATCACC 2636
 QY 229 SerAlaArgSerSerLeuAlaProAspSerPheTrpAlaAsnHisAspSerLysTyr 248
 Db 2635 TCGGCC-----TGAACCTCAAGCGGACTCGCGACC 2606
 QY 249 AlaTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeu 268
 Db 2605 AACACCATCTCCGTAGTCTCGATGGACACCGCGCGCGCGCGCGCGCTC--- 2549
 QY 269 ArgGluHisPheValLysAsnArgGlyIleThrProLys-----ProSerLeuLeu 285
 Db 2548 -----CACCTCGCGCGCAACCCCTCGCGCACCGCTCCCGAGTCCGCGCGCTGACG 2495
 QY 286 LysAlaAlaLeuIleAlaGlyAlaAspValGlyLeuGlyTyrProAsnGlyAsnGln 305
 Db 2494 TCCCGCGCACACCGCGCGTCTCACCACCGCGCGACGGGCTCGGCCAAC----- 2444
 QY 306 GlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyrValAsnGlu----- 323
 Db 2443 -----CGGCTCTCTAGTCTCGCGCGCGCGCAC 2417
 QY 324 -----SerSerAlaLeuSerThrSerGlnLysAlaThrTyrThrPheThrAla 339
 Db 2416 GACCACCTCCGGCGCGCGCTTCGAGACACCGGTGACTACAGTACGACGACCACTCC 2357
 QY 340 ThrAlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThr 359
 Db 2356 ACGGTGAGTCCCGGTGAGCGTCTCCGGCGTCTCCGCGCAACGGCGCTCGGCCCTCGCC 2297
 QY 360 AlaSerVal-----ThrLeuValAsnAspLeuAspLeuValIleThrAlaPro 375
 Db 2296 GTAGAGTTCACATCGTCCACACGTATATCGCGACCTCCAGGTCCAGTATCGCCCCC 2237
 QY 376 AsnGlyThrArgTyrValGlyAsnAspPheSerAlaProPheAsnAsnTrpAspGly 395
 Db 2236 GACGCGCGGTACACGCTCAAGTGTACGGCACC-----GGCGGAGTTCGGACAAC 2183
 QY 396 ArgAsnAsnValGluAsnValPheIleAsnSer---ProGlnSerGlyThrTyrThrIle 414
 Db 2182 ATCAACACACGATCTCGGTGAGCGCTCTCTCGAGGGCGGCCAACGGCAAGTGGAAATG 2123
 QY 415 GluVal 416
 Db 2122 CGGGTG 2117
 RESULT 14
 US-10-090-624-5

```
; Sequence 5, Application US/10090624
; Publication No. US20020132335A1
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA-6
; CURRENT APPLICATION NUMBER: US/10/090,624
; CURRENT FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: 03/445,472
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 5
; LENGTH: 4765
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
; US-10-090-624-5

Alignment Scores:
Pred. No.: 1,9e-23 Length: 4765
Score: 307.00 Matches: 146
Percent Similarity: 38.93% Conservative: 58
Best Local Similarity: 27.86% Mismatches: 171
Query Match: 13.67% Indels: 149
DB: 13 Gaps: 23

US-09-985-689A-2 (1-434) x US-10-090-624-5 (1-4765)

Qy 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 1024 GGCATGGTATGACATTCATATGTCGACTGACCTTGACCTTACGACCTCACCAGCAA 1083
Qy 41 -----SerMetHisGluAlaPheArgGly 48
Db 1084 GTTCCACTTGGCCAGTACACGGTTACTATGATGTTCTGTTTGTAGTACTACTACGGT 1143
Qy 49 LysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAsn ----- 63
Db 1144 CCTCTCAACTACGCTGTTGCAGAAATAGACCTTACCGAGAAATATGACGATATTGGGTGG 1203
Qy 64 AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlyAlaThrAsn 83
Db 1204 GATGTCACGGTCACGGAACCTCAGTACGTGGAACCTGCTGCTTACGACAGCAACAAT 1263
Qy 83 ----- 83
Db 1264 GATGCTGGGATGGCTCAGTATGATCTGTTGGTAATGGGAAGTTCTCAAGACTCTAT 1323
Qy 84 -----LysGlyMetAlaProGlnAlaAsn 91
Db 1324 GGTGGGATTATACGAAGTTACACAGACACCGCTGCAGGCTGTGCTCCAGGTGCCAA 1383
Qy 92 LeuValPheGlnSerIleMetAspSerSerGlyLeuGlyLeuGlyLeuProSerAsnLeu 111
Db 1384 ATAATGCAATAGAGTTCTT---AGGAGTGATGGACGGGTAGCATGTGGGATATTATA 1440
Qy 112 GlnThrLeuPheSerGlnAlaPheSerAlaGlyAlaArgIleHisThrAsnSerTyrGly 131
Db 1441 GAAGGTATG---ACATACGACGACACCCATGGTGCAGACGTTATAGCATGAGTCTCGGT 1497
Qy 132 AlaAlaValAsnGlyAlaTyrThr-----ThrAspSerArgAsn-----ValAspAsp 147
Db 1498 GGA-----AATGCTCCATCTAGATGGTACTGATCCAGAAACGGTTGCTGTGGATGAG 1551
Qy 148 TyrValArgLysAsnAspMetThrIleLeuPheAlaAlaGlyAsnGluGlyProAsnGly 167
1552 CTTACCGAAAGTAGCGTGTATTCTGTAATAGCTGCAGGAATAGAGTCTCTGGCAT 1611
168 GlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsn 187
1612 AACATCGTTGGAAGTCTCGTGTTCACAAAGCAATAAAGTGTGGAGCTGCTGCA--- 1668
188 LeuArgProSerPheGlySerTyrAlaAsp----- 197
1669 GTGCCCATTAAGTTGGAGTTATGTTTCCCAAGCACCTTGGATATCTCTGATTACTATGA 1728
198 -----AsnIleAsnHisValAlaGlnPheSerSerArgGly 209
1729 TTCTATTACTTCCCGCCCTCACAAACGTT---AGAATAGCATCTCTCTCAAGCAGAGGG 1785
210 ProThrLysAspGlyArgIleLysProAspValMetAlaProGlyThrTyrIleLeuSer 229
1786 CCGAGAATAGATGGTGAATTAACCCCAATAGTGGCTCCAGGTACGGAATTTACTCA 1845
230 AlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsnHisAspSerLysTyrAla 249
1846 TCCTCGCGGATGTGGATTGGCGGAGCTGACTTC----- 1878
250 TyrMetGlyGlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArg 269
1879 ---ATGCTCGAAGCTTCGATGGCTACTCCACATGTCAGCGGTGCTGCTCCTCCTATA 1935
270 GluHisPheValLysAsnArgGlyIleThrProLysProSerLeuLeuLysAlaAlaLeu 289
1936 AGCGGG---GCAAGGCGGAGGAATATACTACAATCCAGATATAATTAAGAAGTCTT 1992
290 IleAlaGlyAlaAla-----AspValGlyLeuGlyTyrProAsnGly 303
1993 GAGAGCGGTGCAACCTGCTTGGGAGATCCATATACTGGCGCAGAAAGTACACTGAGCTT 2052
304 AsnGlnGlyTyrGlyArgVal---ThrLeu-----AspLysSerLeuAsnValAlaTy 320
2053 GACCAAGTCTAGTCTGTTTAAAGTACCAAGTCTCTGGGAATCTTAAGCTATAAAC 2112
320 rValAsnGluSerSerAlaLeuSerThrSerGlnLysAlaThrTyrThrPheThrAlaTh 340
2113 GSCACCACTCTCCCAATTTGATCATCTGGGCGACAGCAAGTCT-----ACAGCGAC 2163
340 rAlaGlyLysProLeuLysIleSerLeuValTyrSerAspAlaProAlaSerThrAl 360
2164 T-----TTGCGGAGTACTTGGGTGTGACGTTATAGAGTCTCTAGCGAAG 2211
360 sSerValThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThr---Arg 379
2212 AACT-----CTATACCTGACATTTGTCGAGTGCACATTAAAG 2247
380 TyrValGlyAsnAspPheSerAlaProPheAsp-----AsnAsnTrp----- 393
2248 TACGTAGGGGACACCGAGTACAGAACTTTTGATCTCTATGCACTGACGCAATGATTAA 2307
394 -----AspGly-----ArgAsnAsnValGlu----- 400
2308 CTTTTTGTGAGTGAAGTGAATTCAGAGAACAAATACCGAGTTGTCTCTAGGTGAAA 2367
401 -----AsnValPheIleAsnSerPro 407
2368 TATGATGATAGAGGTCTTTGACCGAGTCTTATGTTGGAAGGATAATCATTTGATGATCCA 2427
408 GlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsnValProValGlyProGlnAsn 427
2428 ACA-----ACGCCAGTTATTAGACGAGATCTTTGAACACAAATGTTATTCCCAGAGA 2481
428 PheSer 429
2482 TTCCT 2487
RESULT 15
US-09-927-827-33
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 3, 2004, 23:40:05 ; Search time 2279.84 Seconds
(without alignments)
5684.703 Million cell updates/sec

Title: US-09-985-689A-2
Perfect score: 2245
Sequence: 1 NDVARGIVKADVAQSSYGLY.....EVOAYNVPEVGPQNFSLAIVN 434

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-O=/cgn2.1/USPTO_spool/US09985689/runat_31032004_161807_4168/app_query.fasta_1.3498
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=ext -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09985689@cgn_1_1_7084/runat_31032004_161807_4168 -NCPU=6 -ICPU=3
-NO_MMMap -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FCAPOP=6
-FCAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:*
1: em_estba.*
2: em_esthum.*
3: em_estin.*
4: em_estmu.*
5: em_estov.*
6: em_estpl.*
7: em_estro.*
8: em_hci.*
9: gb_estci.*
10: gb_est2.*
11: gb_hci.*
12: gb_est3.*
13: gb_est4.*
14: gb_est5.*
15: em_estfun.*
16: em_estom.*
17: em_gss_hum.*
18: em_gss_inv.*
19: em_gss_pln.*
20: em_gss_vrt.*
21: em_gss_fun.*
22: em_gss_mam.*
23: em_gss_mus.*
24: em_gss_pro.*
25: em_gss_rod.*
26: em_gss_phg.*
27: em_gss_vrl.*
28: gb_gssi.*

29: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match | Length | DB | ID | Description |
|------------|-------|-------|--------|----|-----------|---------------------|
| 1 | 249 | 11.1 | 1605 | 13 | BQ622771 | BQ622771 CC Contig |
| 2 | 231.5 | 10.3 | 4198 | 11 | AK029048 | AK029048 Mus muscu |
| 3 | 230 | 10.2 | 640 | 12 | BJ395336 | BJ395336 Mus muscu |
| 4 | 219.5 | 9.8 | 594 | 12 | BJ393752 | BJ393752 BJ393752 |
| 5 | 216.5 | 9.6 | 532 | 29 | TA319G10P | TA319G10P T. brucei |
| 6 | 208.5 | 9.3 | 2141 | 13 | BQ143519 | BQ143519 Contig M |
| 7 | 203 | 9.0 | 771 | 14 | CA320325 | CA320325 UI-M-FW0- |
| 8 | 200 | 8.9 | 508 | 28 | AQ652212 | AQ652212 Sheared D |
| 9 | 199.5 | 8.9 | 601 | 12 | BJ387574 | BJ387574 BJ387574 |
| 10 | 195 | 8.7 | 633 | 12 | BJ369190 | BJ369190 BJ369190 |
| 11 | 195 | 8.7 | 718 | 12 | BI750157 | BI750157 FG02 10G0 |
| 12 | 193.5 | 8.6 | 1002 | 29 | CNS06D6B | AL393417 T3 end of |
| 13 | 188 | 8.4 | 565 | 13 | BQ770462 | BQ770462 UI-M-F10- |
| 14 | 186 | 8.3 | 716 | 28 | BZ893395 | BZ893395 HL2_0177 |
| 15 | 185 | 8.2 | 675 | 14 | CF727824 | CF727824 UI-M-HB0- |
| 16 | 183.5 | 8.2 | 574 | 29 | TA315H10P | AL490202 T. brucei |
| 17 | 183.5 | 8.2 | 614 | 9 | AJ273402 | AJ273402 AJ273402 |
| 18 | 183 | 8.2 | 641 | 12 | BJ393925 | BJ393925 BJ393925 |
| 19 | 178.5 | 8.0 | 650 | 9 | AJ274038 | AJ274038 AJ274038 |
| 20 | 178 | 7.9 | 530 | 29 | CNS010PO | AL153820 Anopheles |
| 21 | 178 | 7.9 | 895 | 13 | BQ216158 | BQ216158 AGENCOURT |
| 22 | 178 | 7.9 | 2121 | 28 | BZ424995 | BZ424995 100023066 |
| 23 | 178 | 7.9 | 3091 | 11 | BC011275 | BC011275 Mus muscu |
| 24 | 177.5 | 7.9 | 681 | 14 | CB690041 | CB690041 CEST-54-B |
| 25 | 177 | 7.9 | 576 | 14 | CD295943 | CD295943 StrPu691. |
| 26 | 177 | 7.9 | 1572 | 11 | AX107161 | AX107161 Zea mays |
| 27 | 176 | 7.8 | 4662 | 11 | BC060627 | BC060627 Mus muscu |
| 28 | 175.5 | 7.8 | 712 | 14 | CD311344 | CD311344 StrPu691. |
| 29 | 173.5 | 7.7 | 580 | 9 | AJ273745 | AJ273745 AJ273745 |
| 30 | 173.5 | 7.7 | 583 | 9 | AJ273947 | AJ273947 AJ273947 |
| 31 | 173.5 | 7.7 | 593 | 9 | AJ273918 | AJ273918 AJ273918 |
| 32 | 173.5 | 7.7 | 601 | 9 | AJ273921 | AJ273921 AJ273921 |
| 33 | 173.5 | 7.7 | 616 | 14 | CF138007 | CF138007 UI-HF-BNO |
| 34 | 171.5 | 7.6 | 601 | 9 | AJ273050 | AJ273050 AJ273050 |
| 35 | 171.5 | 7.6 | 604 | 9 | AJ273185 | AJ273185 AJ273185 |
| 36 | 171 | 7.6 | 449 | 12 | BJ359939 | BJ359939 BJ359939 |
| 37 | 170.5 | 7.6 | 573 | 14 | CA937626 | CA937626 sav42b10. |
| 38 | 170 | 7.6 | 794 | 14 | CF737198 | CF737198 UI-M-HD0- |
| 39 | 169.5 | 7.6 | 609 | 9 | AJ272712 | AJ272712 AJ272712 |
| 40 | 169 | 7.5 | 781 | 13 | BU612128 | BU612128 UI-M-EMO- |
| 41 | 168.5 | 7.5 | 545 | 13 | BU575479 | BU575479 TGESTzyb8 |
| 42 | 168.5 | 7.5 | 610 | 9 | AJ274218 | AJ274218 AJ274218 |
| 43 | 168.5 | 7.5 | 530 | 13 | EX390734 | EX390734 EX390734 |
| 44 | 168 | 7.5 | 539 | 14 | CD306363 | CD306363 StrPu691. |
| 45 | 168 | 7.5 | 831 | 14 | CA510555 | CA510555 UI-R-FW0- |

ALIGNMENTS

RESULT 1
BQ622771
LOCUS
DEFINITION BQ622771 1605 bp mRNA linear EST 01-JUL-2002
CC Contig67 Conidiobolus cornatus ARSEP 512 Conidiobolus coronatus
CDNA, mRNA sequence.
ACCESSION BQ622771.1 GI:21649940
VERSION BQ622771.1
KEYWORDS EST.
SOURCE Conidiobolus coronatus (Delacroixia coronata)
ORGANISM Conidiobolus coronatus
Eukaryota; Fungi; Zygomycota; Zygomycetes; Entomophthorales;
Ancylistaceae; Conidiobolus.
REFERENCE 1 (bases 1 to 1605)

AUTHORS Freimoser, F. M., Screen, S., Hu, G. and St. Leger, R. J.
TITLE EST analysis of genes expressed by the zygomycete pathogen *Conidiobolus coronatus* during optimized secretion of proteins
JOURNAL Unpublished (2002)
COMMENT Contact: Freimoser F. M.
 Department of Entomology
 University of Maryland
 4112 Plant Sciences Building, College Park, MD 20742, USA
 Tel: 301 405 16 13
 Fax: 301 314 92 90
 Email: ff34@mail.umd.edu.
FEATURES Location/Qualifiers
 source 1..1605
 /organism="Conidiobolus coronatus"
 /mol_type="mRNA"
 /strain="ARSEF 512"
 /db_xref="taxon:34488"
 /clone_lib="Conidiobolus coronatus ARSEF 512"
 /notes="vector: Unizap; Conidiobolus coronatus was grown in minimal medium supplemented with Manduca sexta cuticle and peptone for 18 hours. A cDNA library was constructed in the unidirectional Lambda vector Unizap."

ORIGIN
 Alignment Scores:
 Pred. No.: 4,05e-14 Length: 1605
 Score: 249.00 Matches: 86
 Percent Similarity: 43.99% Conservative: 42
 Best Local Similarity: 29.55% Mismatches: 83
 Query Match: 11.09% Indels: 80
 DB: 13 Gaps: 14

US-09-985-689A-2 (1-434) x BQ622771 (1-1605)
 QY 6 GlyIleValIysAlaAspValAlaGlnSerSer-----TyrGlyLeu----- 19
 DB 718 GGTGTTGCTAAGGCTATTGCCGTCACAGCCAATGCTCTGGGGTCTTCTCGTGTGGT 777
 QY 20 -----TyrGlyGlnGly 23
 DB 778 CAACGTGCTAAGCTCGGATCTGCTCTTCTTACAAACCATGCTGATGCTGCTCAAGGT 837
 QY 24 GlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHis 43
 DB 938 GTCAGTCTTTTCGTTTGTAGTACTGGTGTCAATGTCAGCCACCAATGAC----- 885
 QY 44 GluAlaPheArgGlyLysIleThrAlaLeuTyraLeuGlyArgThrAsnAsnAla--- 62
 DB 886 -----TTCGTTGCTGCTGCCACT-----TGGGTACTAACACTGCTGCT 924
 QY 63 -----AsnAspThrAsnGlnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn 79
 DB 925 GGTAGCAACACTGATGCTCGCGTCACGGTCACTCCTGCTGGTACTATTGCTGGT--- 981
 QY 80 GlyAlaThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMet--- 98
 DB 982 -----ACCACCTATGTTGTGCCAAGAGGCTAACATTTGTCGGTTAAGGCTTAGGT 1035
 QY 99 ---AspSerSerGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGln 117
 DB 1036 GATGATGGCTCCGATCATCTCTCGAATTATCTCCGGTATGTGCTGGTGTGTTAAGCAC 1095
 QY 118 AlaPheSerAlaGlyAlaAraGileHisThrAsnSerTrpGlyAlaAlaValAsnGlyAla 137
 DB 1096 -----TCTGCTGCCAAGAAAGTTATCTCTATGATTAGGAGTGGTGAAGACGATGCT 1149
 QY 138 TyrThrThrAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeu 157
 DB 1150 CTTAACACT-----GCTGTTAACAAACGCTGTACAGAGGAGGTGTCACCTGTGTC 1200
 QY 158 PheAlaAlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLys 177
 DB 1201 ---GCTGCTGTTGAAGATAACAGACAGTCTTGTGGTACTCT---CCGCTTCTGCTCCT 1254

| | | | |
|-----------|------|---|------|
| Qy | 178 | AsnAlaIleThrValGlyAlaThrGlu-----AsnLeuArgProSerPheGlySerTyr | 195 |
| Db | 1255 | TCGCGCATTAACCGTGGTGCCCATGATGCAATGATAAAAGGCTTCACTTCTTCACTTC | 1314 |
| Qy | 196 | AlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGlyProThrIysAspGlyArg | 215 |
| Db | 1315 | GGTAGCTGTGTC----- | 1326 |
| Qy | 216 | IleLysProAspValMetAlaProGlyThrTyrIleIleuSerAlaArgSerSerLeuAla | 235 |
| Db | 1327 | -----GATATCTTAGCTCTCTGGTGTCACCAATCTCTCCACC----- | 1362 |
| Qy | 236 | ProAspSerSerPheTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSer | 255 |
| Db | 1363 | -----TGGAAGGAGTCTAACACGCTCCACCACCAACCATCTCTGGTACCTCT | 1407 |
| Qy | 256 | MetAlaThrProIleValAlaGlyAsnValAla | 266 |
| Db | 1408 | ATGGCTTGGCCCTCACATTCGCTGGTTAGTGTCT | 1440 |
| RESULT 2 | | | |
| AK029048 | | 4198 bp mRNA linear HTC 18-SEP-2003 | |
| LOCUS | | Mus musculus 10 days neonate skin cDNA, RIKEN full-length enriched | |
| FEATURES | | library, clone:473248M1 product:membrane-bound transcription | |
| | | factor protease, site 1, full insert sequence. | |
| ACCESSION | | AK029048 | |
| VERSION | | AK029048.1 GI:26325017 | |
| KEYWORDS | | HTC; CAP trapper. | |
| SOURCE | | Mus musculus (house mouse) | |
| ORGANISM | | Mus musculus | |
| REFERENCE | | | |
| AUTHORS | | Carninci, P. and Hayashizaki, Y. | |
| TITLE | | High-efficiency full-length cDNA cloning | |
| JOURNAL | | Meth. Enzymol. 303, 19-44 (1999) | |
| MEDLINE | | 98279253 | |
| PubMed | | 10349636 | |
| REFERENCE | | | |
| AUTHORS | | Carninci, P., Shibata, Y., Hayata, N., Sugahara, Y., Shibata, K., | |
| | | Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y. | |
| TITLE | | Normalization and subtraction of cap-trapper-selected cDNAs to | |
| | | prepare full-length cDNA libraries for rapid discovery of new genes | |
| JOURNAL | | Genome Res. 10 (10), 1617-1630 (2000) | |
| MEDLINE | | 20495374 | |
| PubMed | | 11042159 | |
| REFERENCE | | | |
| AUTHORS | | Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., | |
| | | Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M., | |
| | | Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., | |
| | | Yanamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., | |
| | | Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., | |
| | | Yoneda, Y., Ishikawa, F., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., | |
| | | Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y. | |
| TITLE | | RIKEN integrated sequence analysis (RISA) system--384-format | |
| | | sequencing pipeline with 384 multicapillary sequencer | |
| JOURNAL | | Genome Res. 10 (11), 1757-1771 (2000) | |
| MEDLINE | | 20530913 | |
| PubMed | | 11076861 | |
| REFERENCE | | | |
| AUTHORS | | | |
| TITLE | | THE RIKEN Genome Exploration Research Group Phase II Team and the | |
| | | FANTOM Consortium. | |
| JOURNAL | | Functional annotation of a full-length mouse cDNA collection | |
| REFERENCE | | Nature 409, 685-690 (2001) | |
| AUTHORS | | | |
| TITLE | | The FANTOM Consortium and the RIKEN Genome Exploration Research | |
| | | Group Phase I & II Team. | |
| JOURNAL | | Analysis of the mouse transcriptome based on functional annotation | |
| REFERENCE | | of 60,770 full-length cDNAs | |
| | | Nature 420, 563-573 (2002) | |
| JOURNAL | | | |
| REFERENCE | | 6 (bases 1 to 4198) | |

AUTHORS

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayata, M., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imocani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M., and Hayashizaki, Y.

TITLE
JOURNAL

Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan [E-mail: genome-res@gsc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216]

COMMENT

CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Please visit our web site for further details.

URL: http://genome.gsc.riken.go.jp/
URL: http://fantom.gsc.riken.go.jp/
Location/Qualifiers

FEATURES
source

1. 4198
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:4732484M11"
/db_xref="MGI:1231080"
/db_xref="taxon:10090"
/clone="4732484M11"
/tissue_type="skin"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="10 days neonate"
417..3575

CDS

/note="unnamed protein product; membrane-bound transcription factor protease, site 1 (MGDI[MGI:1927235, GB|NM_019709, evidence: BLASTN, 99%, match=3782]) putative"
/codon_start=1
/protein_id="BAC26263.1"
/db_xref="GI:26325018"
/translation="MKLVSTWLVLLVLLCKRHLGRLGTRALEKAPCSGHLTLK

VEFSTVVEYEVYVAFNGYFTAKARNSFISALKSEVENWRIIPRNPSSDPSDFE
VIQIKKQKAGLLLEDPNFKRVTPOKRVKFSKFAENPIVPCNETRWSQKWSR
PLKASLSLGSFGWHTGRHSRLLRAI PROVAOTLOADVLMQMGYTGANRVAVFD
TGLSEKHPHFNKVKERTNWNERTLDLGLGHGTFAVGIASMRCEGAFDAELHFR
VFTNQVSYTSWFLDAENYALKKMDVLNLSIGGPDMDHFFVDKXWELTANNVIVS
AIGNDPGLTNPAQMDVIGVGIDFEDNIARFSRGMTTWELPGGYGRVKPDLV
TYGAGRVSGVGGKCRALSGTSVASPVVAGAVTLVSTVQKRELNPASVKQALIASA
RLPEVNNFEQHGKLDLLRAYQILSSYKQASLSYIDTECPYMWPCSQPIYYG
GMPTIVNVNLTNGMVGTRIVDKPEWRPFLPQNGDNI EVAFSYSLWPMWPGYLAISI
SVTKKASMEGIAGCHIMTVASPAETELHSGAETHTVKLPKVKIIPTPRSKRVL
WDOVNLVYPGYGRNERNKMDPLDWDGHDVHTNFDMDYHLRSMGYEVEVLGAPP
TCFDATQGTLLLDVSEBEPPEIAKLIRDVDNGLSLVPSDWNTSVMRKVFDE
NRQWMPEDTQGNIPALNELLVSNMGFSDGLYEGERVLANHDMYVASGSAKFPF
DGVTITQTKQGLVEVLEKTA VAVNVPIGLYQIPSEGGRIIVLYGDSNCLDSRQ
KDCFWLLDALIQYTSYGVTPPSLSNGNRQPPSGAGLAPPERMEGNHLHRSYKLEA
HLGDPKPLPACPKLSWAKPQPLNETAPSNLWKHQLSLDLDKLPNFRNRPQV
RPLSPGEGSAMDIPGIMPGRYNQEVGTIPVPAFLGAMVALAFVVFQISKAKSRPKR
RRPRAKRPOLAAQHPARTPSV"

polyA_signal

4180..4185

/note="putative"

4198

/note="putative"

ORIGIN

Alignment Scores:

| Pred. No.: | 1,32e-11 | Length: | 4198 |
|------------------------|----------|---------------|------|
| Score: | 231.50 | Matches: | 122 |
| Percent Similarity: | 39.14% | Conservative: | 51 |
| Best Local Similarity: | 27.60% | Mismatches: | 165 |
| Query Match: | 10.31% | Indels: | 105 |
| DB: | 11 | Gaps: | 22 |

US-09-985-689a-2 (1-434) x AK029048 (1-4198)

| | | | |
|----|------|--|------|
| QY | 8 | VallysAlaAspValAlaGlnSerSerTyrGlyLeuTyrGlyGlnLeuValAla | 27 |
| DB | 1005 | CTGCAGGCAGATGTCTGTGG---CAGATGGGATACACAGGTCTTAATGTACAGATTGCT | 1061 |
| QY | 28 | ValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGluAlaPheArg | 47 |
| DB | 1062 | GTTTGTGATCTGGGCTC-----AGTGAGACATCCGCAATTTAAG | 1103 |
| QY | 48 | GlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAsn----- | 63 |
| DB | 1104 | AAT-----GTGAAGGAGAGAGAACCAACTGGACCAATGAGCGGACCTG | 1145 |
| QY | 64 | ---AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlyAlaThr | 82 |
| DB | 1146 | GATGATGGCTAGGCATGGCACATCTGTTGAGGT---GTGATGCCAGCATGAGGAG | 1202 |
| QY | 83 | AsnLysGlyMetAlaProGlnAlaAsnLeu---ValPheGlnSerIleMetAspSer | 101 |
| DB | 1203 | TGCCAAGGATTTGCTCCAGATGCAGAGTGCACATCTTCAGGGTCTTTACCAACATCAG | 1262 |
| QY | 102 | GlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaPheSerAla | 121 |
| DB | 1263 | -----GTGCTTACACATCTTGGTCTTCTGGATGCCCTTCACTAT | 1301 |
| QY | 122 | GlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaThrThrAsp | 141 |
| DB | 1302 | GCCATCTTAAAGAGATGACGCTTCTCAACCTTAGCATCGGTGGGCGCCGATTCATGAT | 1361 |
| QY | 142 | SerArgAsnValAspAspTyr-----ValArgLysAsnAspMetThrIleLeuPheAla | 159 |
| DB | 1362 | CATCCGTTTGTGACAAAGGTGGGAATTAACAGTAACTAATGTAATGTTGTTCTGCT | 1421 |
| QY | 160 | AlaGlyAsnGlyGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAla | 179 |
| DB | 1422 | ATTGGCAATGATGACCTCTCTATGGCACTCTGAATAACCTGCTGATGATGATGATG | 1481 |
| QY | 180 | IleThrValGlyAlaThrGlnAsnLeuArgProSerPheGlySerTyrAlaAspSerIle | 199 |
| DB | 1482 | ATTGGAGTGGTGGCATTCAC-----TTTGAAGATTAACATC | 1517 |
| QY | 200 | AsnHisValAlaGlnPheSerSerArgGly-----ProThrLysAsp | 213 |
| DB | 1518 | -----GCTCGCTTTTCTCCAGGGAATGACTACCTGGGAATTACAGAGGCTAT | 1568 |
| QY | 214 | GlyArgIleLysProAspValMetAlaProGlyThrTyrIleLeuSerAlaArgSer | 233 |
| DB | 1569 | GGTGCTGTGAAGCTGACATTCCTC-----ACCTATGGTGGTGGAGTGGGGGTTC | 1619 |
| QY | 234 | LeuAlaProAspSerSerPheTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGly | 253 |
| DB | 1620 | GGTGTGAAGGGGGCTGC-----CGTGCACTCTCAGG | 1652 |
| QY | 254 | ThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheVal | 273 |
| DB | 1653 | ACCAGTGTGCTTCCCGAGTGTGCTGGGGCGCTCACCTTGTAGTAGACACATACAG | 1712 |
| QY | 274 | LysAsnArgGlyIleThrProLysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAla | 293 |
| DB | 1713 | AMCGGGAGCTGGTG-----AATCTGCACTGTGAAGCAGCTTTGATAGCTACGCC | 1766 |
| QY | 294 | AlaAspValGlyLeuGlyTyrProAsnGlyAsnGlyTyrGlyArgValThrLeuAsp | 313 |
| DB | 1767 | CG-GAGACTTCTCTGGGGTCAACATGTTTCGAGCAAGGCTATCG----- | 1807 |

```

QY 314 LysSerLeuAsnValAlaTyrValAsnGluSerSerAlaLeuSerThrSerGlnLysAla 333
Db 1808 -----CAAGTTGGATCTGCTGGAGCTTATCAGATCCTCAGCAG--- 1846
QY 334 ThrTyrThrPheThrAlaThrAlaGlyLysProLeuLys----- 346
Db 1847 -----CTAAACCCAGCAGCAGCTGAGTCTAGCTACATCGACCTGACTGA 1894
QY 347 IleSerLeu-ValTrp-----SerAspAlaProAlaSerThrThrAlaSerValThrLe 364
Db 1895 GTGTCCTACATGTGCGCTACTGTCCAGCCTATCTACTATGAGGAAATGCCAACAA 1954
QY 364 uValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrArgTyrValGlyAsnAs 384
Db 1955 CCGTTAAT-----GTACACATCTCAATGCGATGGCGGTCACAGGAGNAT 1999
QY 384 pPheSerAlaProPheAspAsnAsnTrp-----AspGlyAsnAsnVa 399
Db 2000 TGTGGATAGCCT-----GAGTGGCGACCTTATTACCACAGAAATGGAGACAA 2050
QY 399 lGluAsnValPheIleAsnSer-----ProGlnSerGlyThrTyrThrIleGl 415
Db 2051 TGAAGTGGCCTTCTCTACTCTCTCACTGCTGAGTGTGTGGCGCTGTCAGGTTAC 2110
QY 415 uVal 416
Db 2111 CATT 2114

RESULT 3
BU395336 640 bp mRNA linear EST 08-MAR-2002
LOCUS BU395336 Dictyostelium discoideum cDNA library, SF Dictyostelium
DEFINITION BU395336 Dictyostelium discoideum cDNA clone dds38b16 5', mRNA sequence.
ACCESSION BU395336.1 GI:19306422
VERSION 1
KEYWORDS EST.
SOURCE Dictyostelium discoideum
ORGANISM Dictyostelium discoideum
REFERENCE 1 (bases 1 to 640)
AUTHORS Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
TITL Full length cDNA of Dictyostelium discoideum at the slug stage
JOURNAL Unpublished (2002)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
Location/Qualifiers
1. .640
/organism="Dictyostelium discoideum"
/mol_type="mRNA"
/strain="AX4"
/db_xref="taxon:44689"
/clone="dds38b16"
/sex="mat A"
/dev_stage="slug stage"
/clone_lib="Dictyostelium discoideum cDNA library, SF"

FEATURES
source
Pred. No.: 7,19e-13 Length: 640
Score: 230.00 Matches: 68
Percent Similarity: 50.00% Conservative: 38
Best Local Similarity: 32.08% Mismatches: 70
Query Match: 10.24% Indels: 36
DB: 12 Gaps: 10

US-09-985-689A-2 (1-434) x BU395336 (1-640)

```

```

QY 19 LeuTyrGlyClnGlyClnIleValAlaValAlaAAspThrGlyLeuAspThrGlyArg--- 37
Db 22 TTAAGGTAAGGTAAAGGTAGATATTGATATTGCTGATCTGGTTAGTACCGATCTCT 81
QY 38 -----AsnAspSer-----SerMetHisClnuAlaAAspThrGlyArg 48
Db 82 TTCTTTTCAGATTCAAAGTATCCATATCCATTAATCAAGTCAATGAAATCATAGAAA 141
QY 49 LysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAsnAspThrAsnGlyHis 68
Db 142 GTTGTAACT-----TATATTACTTACCATGACATGAAGATTATGTAATGGTCA 192
QY 69 GlyThrHisValAlaGlySerValLeuGlyAsnGly-----Ala 81
Db 193 GGTACACATGTTTGTGCTCTGCAGCAGGTACTCCAGAGGATTTCTTCATGGCGCTATTCA 252
QY 82 ThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSerSer 101
Db 253 TCATTAGTGTCTTCCAACTGACGCAAGATTGCAATTTATGATCTT-----TATCT 306
QY 102 GlyGlyLeuGlyGlyLeu---ProSerAsnLeuGlnThrLeuPheSerGlnAlaPheSer 120
Db 307 GGAAGTTCTGAACCAACACACCCGAGATTACAGTCAATGTACAAACCATTTATGAT 366
QY 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAla-----AlaValAsnGly 136
Db 367 GCAGGTGCAAGAGTACATGTTGCTTGGGGTCTGTATCTTTGCAAGGTATTATGGT 426
QY 137 AlaTyrThrThrAspSerArgAsnValAspAspTyrValArgLys---AsnAspMetThr 155
Db 427 GGTATTTCGATGATGCTGGTGTATGATGATGATGATGATGATGATGATGATGATGAT 486
QY 156 IleLeuPheAlaAlaGlyAsnGlyProAsnGlyGlyThrIleSerAlaProGlyThr 175
Db 487 ATACTAAGAGCTGCTGCTAAT---AACGAGCTATTGTCATCTTTATTATGCTCAAGCA 543
QY 176 AlalysAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySer--- 194
Db 544 GCTAAAAATCAATTACGTTGTTGCTGAGCAACAGCTCATGTAATTTGTCAGAT 603
QY 195 -----TyrAlaAspAsnIleAsn 200
Db 604 CATTGGATATTATGATTTCTCAGATAATGCTAAT 639

RESULT 4
BU393752 594 bp mRNA linear EST 08-MAR-2002
LOCUS BU393752 Dictyostelium discoideum cDNA library, SF Dictyostelium
DEFINITION BU393752 Dictyostelium discoideum cDNA clone dds32b16 5', mRNA sequence.
ACCESSION BU393752
VERSION 1
KEYWORDS EST.
SOURCE Dictyostelium discoideum
ORGANISM Dictyostelium discoideum
REFERENCE 1 (bases 1 to 594)
AUTHORS Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
TITL Full length cDNA of Dictyostelium discoideum at the slug stage
JOURNAL Unpublished (2002)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
Location/Qualifiers
1. .594
/organism="Dictyostelium discoideum"
/mol_type="mRNA"
/strain="AX4"
/db_xref="taxon:44689"
/clone="dds32b16"

FEATURES
source

```



```

/sex="mat A"
/dev_stage="slug stage"
/clone_lib="Dictyostelium discoideum cDNA library, SP"

ORIGIN
Alignment Scores:
Pred. No.: 7,47e-12 Length: 594
Score: 219.50 Matches: 56
Percent Similarity: 51.3% Conservative: 39
Best Local Similarity: 30.2% Mismatches: 61
Query Match: 9.78% Indels: 29
DB: 12 Gaps: 8

US-09-985-689A-2 (1-434) x BU393752 (1-594)

QY 49 LysilleThrAlaLeuTyAlaLeuGlyArgThrAsnAlaAsnAspThrAsnGlyHis 68
DB 21 AAGTTGTAACTATATATACCATCAACAAGCGAGTAGTAGTAAAGTGGATGTCAC 80
QY 69 GlyThrHisValAlaGlySerValLeuGly 81
DB 81 GGTACACATATTTGGTTCTGCACAGGTACTCCAGAGGATTCTTCAGTTAATATYCA 140
QY 82 ThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 101
DB 141 TCATTAGTGTCTTGCACACTGATCAAAAGATTGCATCTTTGATTGGCAAGTGTCA 200
QY 102 GlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaPheSerAla 121
DB 201 TCAAGTTTGACACCT---CCATCGGATTGAAACAATTATATCAACCATTTATGACGCA 257
QY 122 GlyAlaArgIleHisThrAsnSerTrpGlyAla 137
DB 258 GGTGCAAGAGTGCATTGTGATCTTTGGGTTCTGTATCAGTAGAGGGGTATACAGGTAGT 317
QY 138 TyrThrAspSerArgAsnValAspAspTyrValArgLysAsn---AspMetThrIle 156
DB 318 TATTTCATCAGACACTGCTTCAATTGATGATTTCTTTCATCTCATCTCAGATTTTCATCAT 377
QY 157 LeuPheAlaAlaGlyAsnGluGlyProAsnGlyGlyThrIleSer-----AlaProGly 174
DB 378 CTTAGAGCTGCTGGTAAC-----AACGAGCAATACCTATCATCTACTCATCAATCC 428
QY 175 ThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArg----- 189
DB 429 ACTGCAAGAAGTGTATTACCGTTGGTCTCATCAACAATTCATGAAATTTATTTAACT 488
QY 190 -----ProSerPheGlySerTyrAlaAspAsnIleAsn-----HisValAla 203
DB 489 GATGTCCAAATATATAAATTATCAATCATCTGTCGATATAAATCAAGAGTTAATGT 548
QY 204 GlnPheSerSerArg 208
DB 549 GATTTCGATAGCAGA 563

RESULT 5
TA319G10P 532 bp DNA linear GSS 13-DEC-2000
LOCUS T. brucei sheared genomic DNA clone 319g10, forward sequence,
DEFINITION genomic survey sequence.
ACCESSION AL492464
VERSION AL492464.1 GI:11867408
KEYWORDS GSS.
SOURCE Trypanosoma brucei
ORGANISM Trypanosoma brucei
Eukaryota; Eucenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma.
1 (bases 1 to 532)
AUTHORS Hall, N., Bowman, S., Lennard, N.J., Doggett, J., Atkin, R.,
Chillingworth, C., Ormond, D., Harris, B., El-Sayed, N., Hou, L.,
Melville, S.E., Rajandream, M.A. and Barrell, B.G.
Direct Submission
TITLE Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing
JOURNAL

```

project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and nh@sanger.ac.uk

Constructed at the Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from a cloned population of Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared to give a tight size distribution (4 kb). The v + i method used for the library construction is described in detail in Smith, H. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999).

Email: nelsayed@tigr.org

Details of T. brucei sequencing at the Sanger Centre are available at http://www.sanger.ac.uk/Projects/T_brucei/.

FEATURES

Source
1..532
/organism="Trypanosoma brucei"
/mol_type="genomic DNA"
/strain="TREU927"
/db_xref="taxon:5691"
/clone="319g10"

ORIGIN

Alignment Scores:
Pred. No.: 1.25e-11 Length: 532
Score: 216.50 Matches: 60
Percent Similarity: 48.37% Conservative: 29
Best Local Similarity: 32.61% Mismatches: 68
Query Match: 9.64% Indels: 27
DB: 29 Gaps: 6

US-09-985-689A-2 (1-434) x TA319G10P (1-532)

QY 93 ValPheGlnSerIleMetAspSerSerGlyGlyLeuGly----- 105
DB 2 GTAATGCCAAGTATAATGTGTGCGCCAGGGGGGGAAGATTCTTCAGGGGTGGGCTGCC 61
QY 106 -----GlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaPhe 119
DB 62 CATCCAGTCAGAGCTTGCTCCGCCACGAGTACTCAATATTCTGTCGGGTATAT 121
QY 120 SerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaThr 139
DB 122 GCGCTGGAGCCGCTGTCTCTCAACTCGTGGGTTTGTGCTCCCTCCGAGTATCT 181
QY 140 ThrAspSerArgAsnValAspAspTyrValArgLys---AsnAspMetThrIleLeuPhe 158
DB 182 GCTGTGGAAGAAGATATGATGATGTTGCGAGTAGTATGACGATCGGCTACTTCTC 241
QY 159 AlaAlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsn 178
DB 242 TCCACTGGCAACAGTATCCCAAGATGCC-----CTAATGACTCCGTGCTGTAAGAAC 295
QY 179 AlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsn 198
DB 296 GTGATGTGGTGGGGTCACAAAAACGAG-----TTTGCAGCTTCGAAAGAC--- 343
QY 199 IleAsnHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysPro 218
DB 344 -----ATTGTTCTTCTCGTTTCTTCGCATGGTCCCAACATACGACGGTAGGATGAACCC 397
QY 219 AspValMetAlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSer 238
DB 398 GATCTTGTGTCGTCGCGGGAAGAGGTGTGTCTCTCTCTCTCTGCGAAAGCATCAGCT 457
QY 239 SerPheTrpAlaAsnHisAspSerIstYrAlaTyrMetGlyGlyThrSerMetAlaThr 258
DB 458 AAA-----CAATGTAAGTGGTGGCCAGCGGNGTTTCATCGATGGCAACT 502
QY 259 ProIleValAla 262
DB 503 GCGGCGCTCGCG 514

RESULT 6
BQ142519/c
LOCUS
DEFINITION Contig6 Metarhizium anisopliae sf. acridum ARSEF 324 Metarhizium anisopliae var. acridum cDNA, mRNA sequence.
ACCESSION BQ142519
VERSION BQ142519.1 GI:20279578
KEYWORDS EST.
SOURCE Metarhizium anisopliae var. acridum
ORGANISM Metarhizium anisopliae var. acridum
REFERENCE Freimoser, F.M., Screen, S., Baga, S., Hu, G. and St. Leger, R.J.
AUTHORS EST analysis of genes expressed by two different insect pathogenic fungi during optimized secretion of proteins
TITLE Unpublished (2002)
JOURNAL Contact: Freimoser F. M.
COMMENT Department of Entomology
University of Maryland
4112 Plant Sciences Building, College Park, MD 20742, USA
Tel: 301 405 16 13
Fax: 301 314 92 90
Email: ff34@umd.edu.
FEATURES
source
1..2141
/organism="Metarhizium anisopliae var. acridum"
/mol_type="mRNA"
/strain="ARSEF 324"
/db_xref="taxon:92637"
/clone_lib="Metarhizium anisopliae sf. acridum ARSEF 324"
/notes="Vector: Unizap; Metarhizium anisopliae sf. acridum was grown on insect cuticle and chitin for 24 hours. A cDNA library was constructed in the unidirectional Lambda vector Unizap."

ORIGIN

Alignment Scores:
Pred. No.: 9,18e-10 Length: 2141
Score: 208.50 Matches: 80
Percent Similarity: 41.72% Conservative: 41
Best Local Similarity: 27.59% Mismatches: 112
Query Match: 9.29% Indels: 57
DB: 13 Gaps: 13

US-09-985-689A-2 (1-434) x BQ142519 (1-2141)

Qy 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||

1059 GGTGAGGGTACTTCGGTATATATATGACACTGGTGTGAGGTCTCC-----1012

Qy 41 SerMethisGluAlaPheArgGlyLysIleThrAlaLeuTyArgThrAsn 60
Db :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||

1011 -----CACCCCGAGTTGGCGGTGCGCCACTGGCTCAGGAGCTTCATCAACGGTCAA 958

Qy 61 AsnAlaAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
Db :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||

957 AAC---CGTGTGCCCGCCAGCCATGGAGCTCCTCGCGTACTATT-----GGT 910

Qy 81 AlaThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAsp--- 99
Db :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||

909 AGCCGAGCTACGGTGTTCGCAAAATGCCAAGTCTTCTGCTCAAGTCTTGTATGAC 850

Qy 100 ---SerSerGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAla 118
Db :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||

849 CAGGCGCATGGTTCCTACTCCGGTATCATCAGTGGCATGGACTTTGTTGCCAGGACTCC 790

Qy 119 PheSer-----AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsn 135
Db :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||

789 AAGAGTCGTAACTGCCCAATGGCCACATGCTTCCATGAGTCTGGGA-----742

Qy 136 GlyAlaTyThrThrAspSerArgAsnValAspTyrValArgLysAsnAspMetThr 155
Db :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||

741 GGTGGCTACTCGCGGTCCGTCACACAGGGTGGCCCTGCTTTGGTCAGGTCTGGTGTCTTC 682

Qy 156 IleLeuPheAlaAlaGlyAsnGlyProAsnGlyGlyThrIleSerAlaProGlyThr 175
Db :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||

681 CTTCCGTCGCGCTGGCAACGATAACCGGATGCCCAAAACACCTCT---CCGCTTCC 625

Qy 176 AlalysAsnAlaIleThrValGlyAlaThr-----GluAsnLeuArgProSerPheGly 193
Db :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||

624 GAGCCTACTGCTGCACCTGTTGGTGCACATCGGTACAGTACACCGCATCTACCTTTTCC 565

Qy 194 SerTyralaAspAsnIleAsnHisValAlaGlnPheSerSerArgGlyProThrLysAsp 213
Db :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||

564 AACTAC-----559

Qy 214 GlyArgIleLysProAspValMetAlaProGlyThrTyrlleLeuSerAlaArgSerSer 233
Db :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||

558 GGCAGAGT---GTCGATATCTTCGCTCTGGTACCGCATCTCTCCACC-----511

Qy 234 LeuAlaProAspSerSerPheTrpAlaAsnHisAspSerLysTyThrAlaTyMetGlyGly 253
Db :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||

510 -----TGATCAAT-----GGCCGACCAACACACATCTCTGGC 478

Qy 254 ThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheVal 273
Db :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||

477 ACCTCCATGGCTACTCCCATATTGCTGGT-----CTTGCTGCCTACTTCAGT 430

Qy 274 LysAsnArgGlyIleThrProLysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAla 293
Db :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||

429 GCTCTCAGCGCCAGACTAGCCCTCGCGCTCTTTGCCAGAGATCCAGGACACTTCTACC 370

Qy 294 AlaAspValGlyLeuGlyTyThrProAsnGly 303
Db :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||

369 AAGACGTGATCCCGCATGTGCCCGCTGGC 340

RESULT 7
CA320325 771 bp mRNA linear EST 09-JUL-2003
LOCUS UI-M-FW0-cby-d-23-0-UI.r1 NIH BMAP_FW0 Mus musculus cDNA clone
DEFINITION IMAGE:6816072 5', mRNA sequence.
ACCESSION CA320325
VERSION CA320325.1 GI:24538449
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 771)
AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue procurement: Dr. Jim Lin, University of Iowa
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
This clone was contributed by the Brain Molecular Anatomy Project (BMAP)
Seq primer: pYX-5
Location/Qualifiers
1..771
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:6816072"
/tissue_type="whole brain"
/dev_stages="embryo 13.5,14.5,16.5,17.5dpc"

/lab host="DH108 (T1 phage resistant)"
/clone lib="NIH_BMAP_FW0"
/note="Organ: Brain; Vector: pYX-Asc; Site: 1: EcoR I;
Site 2: Not I; The library was constructed according
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured RNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with oligo-dT
primer containing a Not I site. Double strand cDNA was
size selected according to mRNA size fraction, ligated
with EcoR I adaptor, digested with NotI and then cloned
directionally into pYX-Asc vector. The library tag
sequence located between the Not I site and the polyA tail
is ACGGAGACAG. This library was created for the University
Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
Developing Mouse Nervous System', supported by National
Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
program coordinator."

ORIGIN

Alignment Scores:
Pred. No.: 5,696-10 Length: 771
Score: 203.00 Matches: 84
Percent Similarity: 41.81% Conservative: 36
Best Local Similarity: 29.22% Mismatches: 109
Query Match: 9.04% Indels: 58
DB: 14 Gaps: 13

US-09-985-689A-2 (1-434) x CA320325 (1-771)

| | | | |
|----|-----|--|-----|
| Qy | 8 | ValLysAlaAspValAlaGlnSerTyrGlyLeuTyrGlyGlnGlyValAla | 27 |
| Db | 40 | CTGCGAGGAGATGCTGTGG---CAGATGGGATACACAGGTGCTAATGTCAGAGTGTCT | 96 |
| Qy | 28 | ValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGluAlaPheArg | 47 |
| Db | 97 | GTTTGTGATCTAGTGGCTC-----AGTGAGAGCATCCGCAATTTTAAG | 138 |
| Qy | 48 | GlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAsn | 63 |
| Db | 139 | AAT-----GTGAGGAGAGAACCACTGGACCAATGAGCGGACCCCTG | 180 |
| Qy | 64 | ---AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlyAlaThr | 82 |
| Db | 181 | GATGATGGCTAGGCGACATTCGTGAGT---GTGATGCCAGCATGAGGGAG | 237 |
| Qy | 83 | AsnLysGlyMetAlaProGlnAlaAsnLeu---ValPheGlnSerIleMetAspSerSer | 101 |
| Db | 238 | TGCCAAGGATTGCTCCAGATGCAGAGCTGCACATCTTCAGGGTCTTTTACCACCAATCAG | 297 |
| Qy | 102 | GlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaPheSerAla | 121 |
| Db | 298 | -----GTGCTTACACATCTGCTTCTGATGCTTCACTAT | 336 |
| Qy | 122 | GlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyrThrThrAsp | 141 |
| Db | 337 | GCCATCTCTAAAGAGATGACGCTTCTCAACCTTAGCATCGTGGCCGCACTTCATGAT | 396 |
| Qy | 142 | SerArgAsnValAspAspTyr-----ValArgLysAsnAspMetThrIleLeuPheAla | 159 |
| Db | 397 | CATCCGTTGTGACAAGGTGGGAATTAACAGCTACATGTAATATGTTGCTTCTGCT | 456 |
| Qy | 160 | AlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAla | 179 |
| Db | 457 | ATTGGCAATGATGAGACCTCTCTATGGCACTCTGAATACCCCTGCTGATCAGATGCTG | 516 |
| Qy | 180 | IleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIle | 199 |
| Db | 517 | ATTGAGTGGGGGCAATTGAC-----TTTGAAGAATAACATC | 552 |
| Qy | 200 | AsnHisValAlaGlnPheSerSerArgGly-----ProThrLysAsp | 213 |
| Db | 553 | -----GTCGCTTTTCTCCAGGGGAATGACTACTCTGGGAATTACCGAGGGCTAT | 603 |

| | | | |
|----|-----|--|-----|
| Qy | 214 | GlyArgIleLysProAspValMetAlaProGlyThrTyrIleLeuSerAlaArgSerSer | 233 |
| Db | 604 | GGTGTGTGAAGCTGACATTGTC-----ACCTATGGTCTCGAGTGGGGTTC | 654 |
| Qy | 234 | LeuAlaProAspSerSerPheTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGly | 253 |
| Db | 655 | GGTGTGAAGGGGGCTGC-----CGTGCACTCTCAGGG | 687 |
| Qy | 254 | ThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheVal | 273 |
| Db | 688 | ACAGTGTGCTCTCCCAAGTGGTGGCGGTCCCTTGTAGTANGCACAGTACAG | 747 |
| Qy | 274 | LysAsnArgGlyIleThrPro | 280 |
| Db | 748 | AAGCGGGAGCTGCTGATCCT | 768 |

RESULT 8
AQ652212/c
LOCUS
DEFINITION
Sheared DNA-8P2.TF Sheared DNA Trypanosoma brucei genomic clone
Sheared DNA-8P2, genomic survey sequence.
ACCESSION
AQ652212
VERSION
AQ652212.1 GI:5145398
KEYWORDS
GSS.
SOURCE
Trypanosoma brucei
ORGANISM
Trypanosoma brucei
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma.

REFERENCE
1 (bases 1 to 508)
El-Sayed,N., Zhao,S., Zhao,H., Gill,S., Suh,E., Malek,J., Fujii,C.,
Gerrard,C., Leech,V., de Jong,P., Ollu,E., Melville,S.,
Doneelson,J., Fraser,C. and Adams,M.
Determination of clone end sequences from Trypanosoma brucei GUTat
10.1 sheared DNA library
Unpublished (1999)
Other GSSs: Sheared DNA-8P2.TR
Contact: Najib M. El-Sayed
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: nleayed@tigr.org

Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared
DNA library constructed at TIGR. Clones will be available for
distribution through ATCC. Sheared DNA end sequences search page:
http://www.tigr.org/tldb/mdb/tbdb/
Seq primer: M13-Forward
Class: shotgun.
Location/Qualifiers
1 508
/organism="Trypanosoma brucei"
/mol_type="genomic DNA"
/strain="TREU927/4 GUTat 10.1"
/db_xref="taxon:5691"
/clone="Sheared DNA-8P2"
/clone_lib="Sheared DNA"
/notes="Vector: pUC18; Site 1: SmaI. Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from a cloned population of
Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically
sheared to give a tight size distribution (approx 2 kb).
The v + i method used for the library construction is
described in detail in Smith, H.O. and Venter, J.C.
(Making small insert libraries for whole genome shotgun
sequencing projects. In Genome Sequencing: A Practical
Approach, eds. M. Vaudin and B. Barell, Oxford University
Press, 1999)."

FEATURES
source
1 508
/organism="Trypanosoma brucei"
/mol_type="genomic DNA"
/strain="TREU927/4 GUTat 10.1"
/db_xref="taxon:5691"
/clone="Sheared DNA-8P2"
/clone_lib="Sheared DNA"
/notes="Vector: pUC18; Site 1: SmaI. Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from a cloned population of
Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically
sheared to give a tight size distribution (approx 2 kb).
The v + i method used for the library construction is
described in detail in Smith, H.O. and Venter, J.C.
(Making small insert libraries for whole genome shotgun
sequencing projects. In Genome Sequencing: A Practical
Approach, eds. M. Vaudin and B. Barell, Oxford University
Press, 1999)."

ORIGIN
Alignment Scores:
Pred. No.: 5,596-10 Length: 508
Score: 200.00 Matches: 57

```

Percent Similarity: 50.30%      Conservative: 28
Best Local Similarity: 33.73%    Mismatches: 56
Query Match: 8.91%             Indels: 28
DB: 28                         Gaps: 7

US-09-985-689A-2 (1-434) x AQ652212 (1-508)

QY 156 IleLeuPheAlaAlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThr 175
DB 503 CTTATCTTCTCCACTGGCACAGTATCCAGATGGC-----CTAATGACTCCGGTGGT 450
QY 176 AlaLyAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheLySerTyr 195
DB 449 GGTAAAGACGTGATGCGTGGGGTCCACAAAACGTG-----TTTACGCTTGG 399
QY 196 AlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArg 215
DB 398 AAGAC-----ATTGTTCTTCTGTTTCTTCGATGGTCCACATACGCGTAGG 348
QY 216 IleLyProAspValMetAlaProGlyThrTyrIleLeuSerAlaArgSerLeuAla 235
DB 347 ATGAACCCGATCTTGTGGTCCCGGGAAGAGGTGTCTCTCTCTCTCTCTCTCTCT 288
QY 236 ProAspSerSerPheTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyThrSer 255
DB 287 GCATCAGCAAA-----CAATGTAAGTGTGGTGGCCCAAGCGGGTTCATCG 243
QY 256 MetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLys 274
DB 242 ATGCAACTGGCGCGTGGCGCGCGCTACACTGTTGGTCAATGATGTACGAGGCTA 183
QY 275 AsnArgGlyIleThrProLysProSerLeuLeuLysAlaAlaLeuIle----- 290
DB 182 AATCGAACTGCTTCACTTCTGCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 123
QY 291 -----AlaGlyAlaAlaAspValGlyLeuGlyTyr----- 300
DB 122 CCGTTGAGTAATCCACCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 63
QY 301 ProAsnGlyAsnGlnGlyTrpGlyArg 309
DB 62 CCAACGGGACACAGGGATGTTCCGC 36

RESULT 9
LOCUS BJ387574 601 bp mRNA linear EST 08-MAR-2002
DEFINITION BJ387574 Dictyostelium discoideum cDNA library, SF Dictyostelium
          discoideum cDNA clone dds3a18 5', mRNA sequence.
ACCESSION BJ387574.1 GI:19296958
KEYWORDS EST.
SOURCE Dictyostelium discoideum
        Dictyostelium discoideum
        Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE 1 (bases 1 to 601)
AUTHORS Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
TITLE Full length cDNA of Dictyostelium discoideum at the slug stage
JOURNAL Unpublished (2002)
COMMENT Contact: Tadasu Shin-i
        Center For Genetic Resource Information
        National Institute of Genetics
        1111 Yata, Mishima, Shizuoka 411-8540, Japan
        Tel: 81-559-81-6856
        Fax: 81-559-81-6855
        Email: tshini@genes.nig.ac.jp.
        Location/Qualifiers
FEATURES
    source
        1..601
            /organism="Dictyostelium discoideum"
            /mol_type="mRNA"
            /strain="AX4"
            /db_xref="taxon:44699"
            /clone="dds3a18"
            /sex="mat A"

```

```

/dev_stage="slug stage"
/clone_lib="Dictyostelium discoideum cDNA library, SF"

ORIGIN
Alignment Scores:      8.42e-10      Length:      601
Pred. NO.:            199.50         Matches:     59
Score:                44.86%         Conservative: 24
Best Local Similarity: 31.89%         Mismatches:  63
Query Match:          8.09%          Indels:      39
DB:                   12             Gaps:        6

US-09-985-689A-2 (1-434) x BJ387574 (1-601)

QY 198 AsnIleAsnHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgGlyLys 217
DB 51 AATGAGTAATAATTTGTTTCATTCATCAAAAGGTCACACATCATGATGTAGATGAA 110
QY 218 ProAspValMetAlaProGlyThrTyrIleLeuSerAlaArgSerLeuAla----- 235
DB 111 CCTGATTAGTTGGCCCTGGTGAATATATTACATCGGCAGATCAATAATGGTGCAATACA 170
QY 236 -----ProAspSerSerPheTrpAlaAsnHisAspSerLysTyrAlaTyrMet 251
DB 171 ACAGACCAATGTGGTGGTCTTTA---CCAAATACAAATGCATTATTGGCG---ATA 224
QY 252 GlyGlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHis 271
DB 225 TCTGTATCATCAATGGCAACCTCATTTGCGAGCAGCAGCAACAACAATCTTAGACAATAT 284
QY 272 -----PheValLysAsnArgGlyIleThrProLys 281
DB 285 TTAGTTGATGTTATTATCCAACTGGTTCATTTAGAAATCAATAATAATAATTAACAACCACT 344
QY 282 ProSerLeuLysAlaAlaLeuIleAlaGlyAla----- 293
DB 345 GGATCAATTATAAGCATTAATGATTAAATGCTCAGTTATTAAATGGTACATTTCAA 404
QY 294 -----AlaAspValGlyLeuGlyTyrProAsnGlyAsn----- 304
DB 405 TTGATTACATCATCAAGTATTACATATCCATCAACCAAGTTTTTGAAATTTTGCAGGT 464
QY 305 -----GlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
DB 465 GCAAGTTAGTTCAGGTGGGTCCTATTAGATAGTAAATGTTGTTGTTGTTGTTGTTGTT 524
QY 321 ValAsnGluSerSerAlaLeuSerThrSerGlnLysAlaThrTyrThrPheThrAlaThr 340
DB 525 AATAATAATAGTAATAATAATAATAAACAATCAGATGATGATACTAAATTTGTTGTTATT 584
QY 341 AlaGlyLysProLeu 345
DB 585 GGTGGATTAGATTTA 599

RESULT 10
LOCUS BJ369190 633 bp mRNA linear EST 08-MAR-2002
DEFINITION BJ369190 Dictyostelium discoideum cDNA library, CF Dictyostelium
          discoideum cDNA clone ddc49116 5', mRNA sequence.
ACCESSION BJ369190
KEYWORDS EST.
SOURCE Dictyostelium discoideum
        Dictyostelium discoideum
        Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE 1 (bases 1 to 633)
AUTHORS Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
TITLE Full length cDNA of Dictyostelium discoideum at the culmination
JOURNAL Unpublished (2002)
COMMENT Contact: Tadasu Shin-i
        Center For Genetic Resource Information
        National Institute of Genetics

```

1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.

FEATURES

Location/Qualifiers
1..633
/organism="Dictyostelium discoideum"
/mol_type="rRNA"
/strain="AX4"
/db_xref="taxon:44689"
/clone="ddc49116"
/sex="mat A"
/dev_stage="Culmination stage"
/clone_lib="Dictyostelium discoideum cDNA library, CF"

ORIGIN

Alignment Scores:
Pred. No.: 2,666-09 Length: 633
Score: 195.00 Matches: 58
Percent Similarity: 41.55% Conservative: 28
Best Local Similarity: 28.02% Mismatches: 49
Query Match: 8.69% Indels: 72
DB: 12 Gaps: 6

US-09-985-689A-2 (1-434) x BJ369190 (1-633)

QY 136 GlyAlaTyrThrThrAspSerArgAsnValaAspTyrValaArgLys---AsnAspMet 154
DB 6 GTGGTTATTCGATGATGCTGGTGGTATGATGATTCCTCTATGAGTACCCAGAAATC 65
QY 155 ThrLeuPheAlaAlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGly 174
DB 66 TCTATACTAAGAGCTGCTGGTAAT---AAGGAGCTATTGGCATCTTTATTAGTCAAGCA 122
QY 175 ThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySer 194
DB 123 ACAGCTAAATGCAATACATGATGCTGGTGTGAGCAACAGCTCATGTAAATATTGTGPCA 182
QY 195 -----TyrAlaAspAsnIle----- 199
DB 183 GATGATTGGAATATTGATGATTCCTCAGATATGCTAATTTTCAAGACCATGTTATTC 242
QY 199 ----- 199
DB 243 GATAAGAAGTATTGTAATATTATACGACCGCTAAATGTTGTCAGAGGTTTCAATGTTAAA 302
QY 199 ----- 199
DB 303 GGTTCACATATTATGTTGCCAGCATCTATTAAACAAATGCGATCGATTCATTCAACA 362
QY 200 -----AsnHisValAlaGlnPheSerSerArgGlyProThrLys 212
DB 363 CAACTCAATTTTATAATGAAATAATATGCGCATCTTCATCAAGGGTCCACACAT 422
QY 213 AspGlyArgIleLysProAspValMetAlaProGlyThrTyrIleLeuSerAlaArgSer 232
DB 423 GATGGTAGATTGAACTGATATAGTTGCACTGTTGATATATTATACATCGGCAAGATCA 482
QY 233 Ser-----LeuAlaProAspSerSerPheTyrAlaAsnHisAspSer 246
DB 483 AATGGTGAGAAATTCACAGACCAATGTTGGTGTATGTTCTTTA-----CCAAATGCAAT 536
QY 247 LysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGlyAsnVala 266
DB 537 GGTTCTAATGTCATATCTGGTACATCATGCGCAACACCATTTGGCAACAGCAGCAACA 596
QY 267 GlnLeuArgGluHisPheVal 273
DB 597 ATTCTTAGACAATATTAGTT 617

RESULT 11

BI750157

LOCUS

BI750157 718 bp mRNA linear EST 25-SEP-2001

DEFINITION

Fg02_10g08_R_Fg02_AAFc_ECORC_Fusarium graminearum mycelium
Gibberella zeae cDNA clone Fg02_10g08, mRNA sequence.

ACCESSION

BI750157

VERSION

BI750157.1 GI:15771959

KEYWORDS

EST

SOURCE

Gibberella zeae

ORGANISM

Gibberella zeae
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Hypocreomycetidae; Hypocreales; Nectriaceae; Gibberella.

REFERENCE

1 (bases 1 to 718)

AUTHORS

Harris,L.J., Glassco,T., Rocheleau,H., Allard,S., Chapados,J.,
Couroux,P., De Moors,A., Hattori,J.I., Ouellet,T., Robert,L.S.,
Singh,J.A., Sprott,D. and Tinker,N.A.
Expressed Sequence Tags from Fusarium graminearum mycelium
Unpublished (2001)

TITLE

Contact: Harris, Linda J.

JOURNAL

Comment

COMMENT

Eastern Cereal and Oilseed Research Centre
Agriculture and Agri-food Canada
Bldg. 21, Central Experimental Farm, Ottawa, Ontario, K1A 0C6,
CANADA

Tel: (613) 759-1314

Fax: (613) 759-6586

Email: harris@em.agr.ca.

FEATURES

Location/Qualifiers

source

1..718
/organism="Gibberella zeae"

/mol_type="rRNA"

/strain="DAOM 180378"

/db_xref="taxon:5518"

/clone="Fg02_10g08"

/tissue_type="Mycelial tissue"

/dev_stage="Assexual"

/lab_host="E. coli (Sure cells)"

/note="Vector: Bluescript SK+/XhoI-ECOR1; Site_1: EcoRI;
Site_2: XhoI; Mycelial tissue was collected from V8 agar
plates after a growth period of 6-7 days at 25 C with 14
hrs (FL/UV) day light exposure. Mycelia was ground in
liquid nitrogen prior it's storage at -80 C until RNA
extraction. Directional cloning with 5' end of cDNA cloned
into EcoRI site of pBluescript and 3' end of cDNA cloned
into XhoI site of pBluescript (Stratagene, La Jolla, CA)."

ORIGIN

Alignment Scores:

Pred. No.: 3,3e-09 Length: 718
Score: 195.00 Matches: 67
Percent Similarity: 43.69% Conservative: 30
Best Local Similarity: 30.18% Mismatches: 89
Query Match: 8.69% Indels: 36
DB: 12 Gaps: 10

US-09-985-689A-2 (1-434) x BI750157 (1-718)

QY 46 PheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAsnAspThr 65

DB 9 TTCAGGGTCGTCTCAGGCTGTCTACACTGCTTCAGCGCCAGACCGCT---GACACC 65

QY 66 AsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn-----GlyAlaThrAsn 83

DB 66 AACGGTCACGGAACCTCAGCTTGTCTGCACTATTTCGCGAAAGACATACGTTGTGCCAAG 125

QY 84 LysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSerSerGlyGly 103

DB 126 AAGGCCACCATC---CAAGCTGTCAAGGTTCTCCAGGT-----AGTTCAATCCAGC 173

QY 104 LeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaPheSerAlaGlyAla 123

DB 174 ACCTCCATCATCTCGTGGCTTCAACTGGGCTGCAACAGACATCATCTCCAAGGGC--- 230

QY 124 ArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyrThrAspSerArg 143

DB 231 CGAACCAAGACACTCAGTCGTCATATGTCTCTCGCGGGTGGTACTCTGCTTCTCTCAAC 290

Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.

FEATURES

```

source
    1. 1002
    /organism="Zygosaccharomyces rouxii"
    /mol_type="genomic DNA"
    /strain="CBS 732"
    /db_xref="taxon:4956"
    /clone="AR0A010H04"
    /clone_lib="AR0AA"
    /notes="end : T3"
    <2..>979
    /misc_feature
    /notes="similar to Saccharomyces cerevisiae ORF YCR045c [
    /evidence-not_experimental

ORIGIN
Alignment Scores:      8.38e-09      Length:      1002
Pred. No.:            193.50      Matches:      80
Score:                41.90%      Conservative:  39
Percent Similarity:   28.17%      Mismatches:   102
Best Local Similarity: 8.62%      Indels:       63
Query Match:         29          Gaps:        14
DB:

US-09-985-689A-2 (1-434) x CNS06D6B (1-1002)

QY      14  GlnSerSerTyrGlyLeuTyr-----GlyGlnGlyGlnIleValala 27
      278  GAAGAGAGTTTCATTATTTACTATTATAATGGCACAAGGGTAGAAATATCAATGCATAT 337
      28  ValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGluAlaPheArg 47
      338  ATWATTGATCTACTGGAATC-----TATAAGAACACATAAGGACTTCTGT 379
      48  GlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAla---AsnAspThrAsn 66
      390  GGAGGT-----GCATATTGGCGCGAGACTTTACAGGGGAAGGTCCTGGTGATCGTAAT 433
      67  GlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlyAlaThrAsnLysGlyMet 86
      434  GGACATGGAATCAGTTCGTTGCGCATTTGTG-----GGATCGAGCAACTTTGGTGT 484
      87  AlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSerSerGlyGlyLeuGlyGly 106
      485  GCTAAAGAGTGAATTTAATTGACGTTAAAGCTTTGACCAATAGAGGC----- 532
      107  LeuProSerAsnLeuGlnThrLeuPheSer-----GlnAlaPhe 119
      533  ---CAAGGCAATTTAACACACAGTAATTAGCGCTCTTGAATTTGCCGTGAATTCATTGTAAA 589
      120  SerAlaGlyAlaArg-----IleHisThrAsnSerTyrGlyAlaAlaValAsnGlyAla 137
      590  TCATCGGCAAAAAGGGTGGTGGTCCCATCTTTTCATTTGGTGGCGCTCAGAACAGTGTA 649
      138  TyrThrThrAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeu 157
      650  ATAATATCAAGCAATTTAAA-----GCTGCTCATGAAGCAGCGGCTCATTTATTGTA 697
      158  PheAlaAlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLys 177
      698  GTTGAGCAGGTAATTCCTAATTAATTAAT---GCATGTGGGAATAGCCCTGCCCTCGCACC 754
      178  AsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAsp 197
      755  GAGGCTATTACTTAGGGGCC-----TTTGTGAT 784
      198  AsnIleAsnHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLys 217
      785  CGTACAGATACAAATTTGCAAAATTTAGTAATTTGGGGTCCATGT----- 826

```

```

QY      144  AsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAlaGlyAsnGlu 163
      291  AACGGCTGTCGAGTCTKCYTCCAGCTCCGGTATTATCTGTGCATTCGTCGGGTAAACGAT 350
      164  GlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAlaAlaIleThrValGly 183
      351  GGTGCCAACGCTGCCAACACTTCT---CTGCTCTTCTCCAGCGCCAWCACTGTCGT 407
      184  AlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsnHisValala 203
      408  GCCATTGACAGC-----AACTGGGCC-----ATTGCC 434
      204  GlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspValMetAlaPro 223
      435  TCGTACTCAACTACGCTACCGTCTCTC-----GATATCTTCGYCCT 476
      224  GlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheThrAlaAsn 243
      477  KGCACACAGCGTTCTCTCCGCC-----TGGTACACC 506
      244  HisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGly 263
      507  AGCAACAGTCCACCAACACCATCAGCGKALSTCCTGCTACTCYCCACATTCGCCGA 566
      264  AsnVal 265
      567  CTTGTC 572

```

```

RESULT 12
CNS06D6B
LOCUS
DEFINITION
T3 end of clone AR0A010H04 of library AR0AA from strain CBS 732 of
Zygosaccharomyces rouxii, genomic survey sequence.
ACCESSION
AL393417
VERSION
AL393417.1 GI:12143597
KEYWORDS
GSS.
SOURCE
Zygosaccharomyces rouxii
ORGANISM
Zygosaccharomyces rouxii
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Zygosaccharomycetes.
REFERENCE
1 (bases 1 to 1002)
AUTHORS
Souciet J.L., Aigle M., Artiguenave F., Blandin G.,
Bollotin-Fukuhara M., Bon E., Brottier P., Casaregola S.,
de Montigny J., Dujon B., Durrien P., Lepingle A., Llorente B.,
Maupertuy A., Neuveglise C., Ozier-Kalogeropoulos O., Potier S.,
Saurin W., Tekalia F., Toffano-Nioche C., Wesolowski-Louvel M.,
Wincker P. and Weissenbach J.
Genomic exploration of the hemiascomycetous yeasts: 1. A set of
yeast species for molecular evolution studies
GENOMIC EXPLO. OF THE HEMIASCOMYCETOUS YEASTS 1: A SET OF
YEAST SPECIES FOR MOLECULAR EVOLUTION STUDIES
FEMS Lett. 487 (1), 3-12 (2000)
1152876
PUBMED
20584711
JOURNAL
MEDLINE
TITLE
2 (bases 1 to 1002)
de Montigny J., Straub M., Potier S., Tekalia F., Dujon B.,
Wincker P., Artiguenave F. and Souciet J.
Genomic exploration of the hemiascomycetous yeasts: 8.
Zygosaccharomyces rouxii
FEMS Lett. 487 (1), 52-55 (2000)
20584718
PUBMED
1152883
JOURNAL
MEDLINE
TITLE
3 (bases 1 to 1002)
Genoscope.
Direct Submission
Submitted (06-SEP-2000) Genoscope - Centre National de Sequencage,
2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
secre@genoscope.cns.fr - Web : www.genoscope.cns.fr)
This GSS is part of a random genomic sequencing program of thirteen
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
exiguus, Saccharomyces servazii, Zygosaccharomyces rouxii,
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,

```

QY 218 ProaspValMetAlaProGlyThrTyrlleLeuSerAlaArgSerSerLeuAlaProasp 237
 Db 827 GTGATATTTCATCGATCAGCGCTRAAGTAAATCATTCG---TCGCATTCCCACT--- 880
 QY 238 SerSerPheTrpAlaAsnHisaspSerlystYrAlaTyMetGlyGlyThrSerMetAla 257
 Db 881 -----CATAAACCAATTGCATTTCAGGTACTTCGATGCT 916
 QY 258 ThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheVallyAsnArgGly 277
 Db 917 TCACCAAGGTTACTGACTTCTGCTATATTATTGGATAAGGCGTTCAACCCAGAAAT 976
 QY 278 IleThrProLys 281
 Db 977 ATTAAAGCTAAG 988

RESULT 13
 BQ770462
 LOCUS
 DEFINITION
 UI-M-F10-bvV-m-19-0-UI.r1 NIH BMAP_F10 Mus musculus cDNA clone
 IMAGE:5702970 5', mRNA sequence.

ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Mus musculus (house mouse)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 665)
 NIH-MGC http://mgs.nsl.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgaabs@emil.nih.gov

Tissue Procurement: Dr. Jim Lin, University of Iowa
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 This clone was contributed by the Brain Molecular Anatomy Project
 (BMAP)

Seq primer: pYX-5.
 Location/Qualifiers
 1. .665
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /clone="IMAGE:5702970"
 /tissue_type="whole brain"
 /dev_stage="embryo 12.5dpc"
 /lab_host="DH10B (T1 phage resistant)"
 /clone_lib="NIH BMAP F10"
 /notes="Organ: Brain; Vector: pYX-Asc; Site 1: EcoR I;
 Site 2: Not I; The library was constructed according
 Benaldo, Lennon and Soares, Genome Research, 6:791-806,
 1996. Denatured RNA was size fractionated on a 1% agarose
 gel. First strand cDNA synthesis was primed with oligo-dT
 primer containing a Not I site. Double strand cDNA was
 size selected according to mRNA size fraction, ligated
 with EcoR I adaptor, digested with NotI and then cloned
 directionally into pYX-Asc vector. The library tag
 sequence located between the Not I site and the polyA tail
 is GAGCCACGAC. This library was created for the University
 of Iowa Brain Anatomy Project (BMAP). 'Gene Discovery in the
 Developing Mouse Nervous System', supported by National
 Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
 program coordinator."

ORIGIN

Alignment Scores:

Pred. No.: 1.5e-08 Length: 665
 Score: 188.00 Matches: 71
 Percent Similarity: 44.58% Conservative: 36
 Best Local Similarity: 29.58% Mismatches: 91
 Query Match: 8.37% Indels: 42
 DB: 13 Gaps: 10
 US-09-985-689A-2 (1-434) x BQ770462 (1-665)
 QY 63 AsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlyAlaThr 82
 Db 14 GATGATGGGTAGCCATGCGCATTGCTTGCAGGT---GTGATTCCAGCATGAGGGAG 70
 QY 83 AsnLysGlyMetAlaProGlnAlaAsnLeu---ValPheGlnSerIleMetAspSerSer 101
 Db 71 TGCCAGGATTGCTCCAGATGCGAGGTCACATCTTCAGGCTCTTACCAACATCAG 130
 QY 102 GlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaPheSerAla 121
 Db 131 -----GTGCTCTACATCTTGGTTTCTGGATCGCTTCAACTAT 169
 QY 122 GlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyThrThrAsp 141
 Db 170 GCATCTCTAAAGAGATGGACGTTCTCACTTAGCATCGTGCGGCCGACCTTCAGGAT 229
 QY 142 SerArgAsnValAspAspTyr-----ValArgLysAsnAspMetThrIleLeuPheAla 159
 Db 230 CATCCGTTTCTGACAGGTTGGGAATTAACAGCTAACAAATGTAATATTATGTTCTGCT 289
 QY 160 AlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAla 179
 Db 290 ATTGGCAATGATGACCTCTCTATGGCATTCTGTAACCTCTGATCATGATGATGATG 349
 QY 180 IleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIle 199
 Db 350 ATTGGAGTGGTGGCATTGAC-----TTTGAAGATAAACATC 385
 QY 200 AsnHisValAlaGlnPheSerSerArgGly-----ProThrLysAsp 213
 Db 386 -----CCTCGCTTTCTCCAGGGGAATGACTACTCTGGGAATTAACAGGAGCTAT 436
 QY 214 GlyArgIleLysProAspValMetAlaProGlyThrTyrlleLeuSerAlaArgSerSer 233
 Db 437 GGTCTGTGAAGCTGACATTGTC-----ACCTATGGTCTGGAGTGGGGGTCTCC 487
 QY 234 LeuAlaProAspSerSerPheTrpAlaAsnHisAspSerLysTyrAlaTyMetGlyGly 253
 Db 488 GGTGTGAAGGGGGCTGC-----CGTCACACTCTCAGGG 520
 QY 254 ThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheVal 273
 Db 521 ACCAGTGTCTTCNCAGTGGTGGTGGCGCGTCACTTGTGTAGTAAAGCACAGTACAG 580
 QY 274 LysAsnArgGlyIleThrProLysProSerLeuLysAlaAlaLeuIleAlaGlyAla 293
 Db 581 AAGCGGAGTGGTGT-----AATCTGCGCAGTGTGAAGCACGCTTTGTAGCGTCAGCC 634
 RESULT 14
 BZ893395/c
 LOCUS
 DEFINITION
 H12.0177 H1 pUC18 Library Halorubrum lacusprofundi genomic 5',
 genomic survey sequence.
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Halorubrum lacusprofundi
 Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
 Halobacteriaceae; Halorubrum.
 1 (bases 1 to 716)
 Gao, Y., Roach, J., Glusman, G., Bhaliga, N.S., Deutsch, K., Pan, M.,
 DasSarma, S., Ng, W. V. and Hood, L.
 Low-pass Sequencing for Microbial Comparative Genomics
 TITLE

3


```
QY 102 GlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaPheSerAla 121
Db 142 -----GTGCTTTACACATCTTGGTTTTCGGATGCCTTCAACTAT 180
QY 122 GlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyThrThrAsp 141
Db 181 GCCATCCTAAGAAGATGGACGTTCTCAACCTTAGCATCGGTGGGCCGACCTTCATGGAT 240
QY 142 SerArgAsnValAspAspTyr-----ValArgLysAsnAspMetThrIleLeuPheAla 159
Db 241 CATCGTTTGTTCACAAAGGTGGGAATTAAACAGCTAACAACTAATATGTTTCTGCT 300
QY 160 AlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAla 179
Db 301 ATTGGCAATGATGGACCTCTCTATGGCACTCTGAATACCCCTGCTGATCAGATGGATGTG 360
QY 180 IleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyAlaAspAsnIle 199
Db 361 ATTGGAGTGGTGGCATTGAC-----TTTGAAGATAACATC 396
QY 200 AsnHisValAlaGlnPheSerSerArgGly-----ProThrLysAsp 213
Db 397 -----GCTCGCTTTTCTCCAGGGGAATGACTACCTGGGAATTACAGGAGCTAT 447
QY 214 GlyArgIleLysProAspValMetAlaProGlyThrTyIleLeuSerAlaArgSerSer 233
Db 448 GGTGCTGTGAAGCCTCACATTGTC-----ACCTATGGTGTGGAGTGGGGGTTC 498
QY 234 LeuAlaProAspSerSerPheThrAlaAsnHisAspSerLysTyAlaTyMetGlyGly 253
Db 499 GGTGTGAAGGGGGCTGC-----CGTGCACCTCTCAGGG 531
QY 254 ThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheVal 273
Db 532 ACCAGTGTGCTTCCCGAGTGTGCTGGGGCGTCCACCTTGTAGTAAGCACACGTACAG 591
QY 274 LysAsnArgGlyIleThrProLysProSerLeuLeuLysAlaAlaLeuIleAla 291
Db 592 AAGCGGAGCTGGTG-----AATCCTGCCAGTGTGAAGCAAGCTTTGATAGCG 639
```

Search completed: April 4, 2004, 11:54:37
Job time : 2290.84 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 3, 2004, 16:47:19 ; Search time 3544.27 Seconds
(without alignments)
5295.174 Million cell updates/sec

Title: US-09-985-689A-3
Perform score: 2263
Sequence: 1 NDVARGIVKADVAQNNGLY.....EVQAINVPSGQRFSLAIVH 433

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters: -MODEL=frame+p2n.model -DEV=xlh
-O=/cgn2_1/USPTO.spool/US09985689/runat_31032004_161806_4155/app_query.fasta_1.3498
-DB=GenEmbl -QEXT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09985689 @CNC 1.1 11925 @runat_31032004_161806_4155 -NCPU=6 -ICPU=3
-NO MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXI=7

Database :

GenEmbl :
1: gb_ba :
2: gb_htg :
3: gb_in :
4: gb_om :
5: gb_ov :
6: gb_pat :
7: gb_ph :
8: gb_pl :
9: gb_pr :
10: gb_to :
11: gb_sts :
12: gb_sy :
13: gb_un :
14: gb_vi :
15: em_ba :
16: em_fun :
17: em_hum :
18: em_in :
19: em_mu :
20: em_on :
21: em_or :
22: em_ov :
23: em_pat :
24: em_ph :
25: em_pi :
26: em_ro :
27: em_sts :
28: em_un :

29: em_vi :
30: em_htg_hum :
31: em_htg_inv :
32: em_htg_other :
33: em_htg_mus :
34: em_htg_pin :
35: em_htg_rod :
36: em_htg_mam :
37: em_htg_vrt :
38: em_sy :
39: em_htg_hum :
40: em_htg_mus :
41: em_htg_other :

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | DB ID | Description |
|------------|--------|---------------|--------|-------|----------------------|
| 1 | 2263 | 100.0 | 1299 | 1 | AB046402 Bacillus |
| 2 | 2251 | 99.5 | 1299 | 1 | AB046405 Bacillus |
| 3 | 2239 | 98.9 | 2218 | 6 | E03808 DNA encodin |
| 4 | 2234 | 98.7 | 1299 | 1 | AB046404 Bacillus |
| 5 | 2040 | 90.1 | 3003 | 6 | AR069954 Sequence |
| 6 | 2040 | 90.1 | 3003 | 6 | BD062155 Nucleic a |
| 7 | 2020.5 | 89.3 | 1302 | 1 | AB046406 Bacillus |
| 8 | 1999.5 | 88.4 | 1923 | 1 | AB084155 Bacillus |
| 9 | 1998.5 | 88.3 | 1920 | 1 | AB046403 Bacillus |
| 10 | 1998.5 | 88.3 | 1920 | 6 | AR368116 Sequence |
| 11 | 1995.5 | 88.2 | 1923 | 6 | AR368118 Sequence |
| 12 | 1894.5 | 88.1 | 1305 | 6 | AX829476 Sequence |
| 13 | 1894.5 | 88.1 | 1823 | 1 | AB051423 Bacillus |
| 14 | 1894.5 | 88.1 | 1823 | 6 | AR368117 Sequence |
| C 15 | 612.5 | 27.1 | 60006 | 1 | AF268611 Unculture |
| C 16 | 559.5 | 24.7 | 301205 | 1 | AE017218 Geobacter |
| 17 | 502 | 22.2 | 5890 | 3 | UDU60086 Dictyosteli |
| 18 | 468 | 20.7 | 5772 | 3 | AF466309 Dictyosteli |
| 19 | 463.5 | 20.5 | 6115 | 3 | DDU20432 Dictyosteli |
| 20 | 443 | 19.6 | 1977 | 6 | AR201152 Sequence |
| 21 | 442.5 | 19.6 | 5109 | 3 | AF263455 Dictyosteli |
| 22 | 422.5 | 18.7 | 1236 | 6 | AR201146 Sequence |
| 23 | 422.5 | 18.7 | 1962 | 6 | AR201155 Sequence |
| C 24 | 422.5 | 18.7 | 12452 | 1 | AE010265 Pyrococcus |
| C 25 | 405 | 17.9 | 126928 | 2 | AC096673 Trypanoso |
| C 26 | 399.5 | 17.7 | 14376 | 1 | AE013049 Thermoana |
| C 27 | 396.5 | 17.5 | 303450 | 1 | SC0939130 Streptomy |
| C 28 | 377.5 | 16.7 | 299850 | 1 | AP004601 Oceanobac |
| 29 | 348.5 | 15.4 | 2539 | 6 | AR202322 Sequence |
| 30 | 347.5 | 15.4 | 1239 | 1 | AF305633 Thermoana |
| 31 | 347.5 | 15.4 | 1239 | 1 | AY028704 Thermoana |
| 32 | 346.5 | 15.3 | 135638 | 1 | AF484556 Streptomy |
| 33 | 345.5 | 15.3 | 3348 | 1 | D83672 Streptomyce |
| C 34 | 345.5 | 15.3 | 12295 | 1 | AE013026 Thermoana |
| 35 | 344 | 15.2 | 292550 | 1 | AP001513 Bacillus |
| 36 | 343.5 | 15.2 | 2809 | 6 | AR202321 Sequence |
| 37 | 343.5 | 15.2 | 3900 | 1 | AB007809 Streptomy |
| 38 | 336 | 14.8 | 1329 | 6 | AX433519 Sequence |
| C 39 | 336 | 14.8 | 300800 | 1 | SC0939112 Streptomy |
| 40 | 334 | 14.8 | 2731 | 1 | AF015225 Thermococ |
| 41 | 329 | 14.5 | 6854 | 1 | BSU39230 Bacillus sp |
| 42 | 329 | 14.5 | 300425 | 1 | AP005044 Streptomy |
| 43 | 327 | 14.4 | 291000 | 1 | SC0939105 Streptomy |
| C 44 | 327 | 14.4 | 302300 | 1 | AP005034 Streptomy |
| C 45 | 321 | 14.2 | 301150 | 1 | AP004602 Oceanobac |

ALIGNMENTS

RESULT 1

```

AB046402      1299 bp      DNA      linear      BCT 16-JUN-2001
LOCUS      Bacillus sp. D6 PROA gene for protease, partial cds.
DEFINITION      AB046402
ACCESSION      AB046402
VERSION      AB046402.1 GI:12381936
KEYWORDS
SOURCE
ORGANISM      Bacillus sp. D6
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
REFERENCE
AUTHORS      1 (sites)
TITLE      Saeki, K., Okuda, M., Hatada, Y., Kobayashi, T., Ito, S., Takami, H. and
            Horikoshi, K.
            Novel oxidatively stable subtilisin-like serine proteases from
            alkaliphilic Bacillus sp.: enzymatic properties, sequences, and
            evolutionary relationships
            Biochem. Biophys. Res. Commun. 279 (2), 313-319 (2000)
JOURNAL
MEDLINE      20568675
PUBMED      11118284
REFERENCE      2 (bases 1 to 1299)
AUTHORS      Saeki, K.
TITLE      Direct Submission
JOURNAL      Submitted (20-JUL-2000) Katsuhisa Saeki, Kao corporation,
            Biological Science Laboratory, Ichikaimachi Akabane 2606, Haga,
            Tochigi 321-3497, Japan (E-mail:1387185@kasanet.kao.co.jp,
            Tel:81-285-68-7400, Fax:81-285-68-7403)
            Location/Qualifiers
FEATURES
source
1..1299
    /organism="Bacillus sp. D6"
    /mol_type="genomic DNA"
    /strain="D6"
    /db_xref="taxon:127889"
gene
1..1299
    /gene="PROA"
    /db_xref="taxon:127889"
CDS
1..1299
    /gene="PROA"
    /codon_start=1
    /transl_table=11
    /product="protease"
    /protein_id="BAB21265.1"
    /db_xref="GI:12381937"
    /translation="NDVARGIVKADVAQNNYVGGQGVAVADTGLDTRNDMSMHE
    AFRKITALYALGRNTNADPNHGHTHVAGSVLGNALNKGAPQANLVFQSIIPSSGG
    LGPLPSNLTSAPGTAKNALTIVGATENYRPSFGSIADNPNIHIAQFSSRGATDRDKPD
    VTAGPFIILASRSLAPDSFVWNVKSYAVNGSTPMATPIVAGNVAILRHFILKVRG
    ITPKELIAALAGTADVGLGTPSGDQNGRVTLDKSLNVAYNNEALITGQKATY
    SFQIQAGKPLKISLVMTDAPGSTTASYTLVNDLDDLTITAPNGQKTVGNDFSYPYDNNW
    DGRNNVENVFVINAPOSQGTITIEVQAYNVESGPQFSLAIVH"
ORIGIN
Alignment Scores:
Pred. No.:      3 09e-134      Length:      1299
Score:      2263.00      Matches:      433
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:      100.00%      Indels: 0
DB:      1      Gaps: 0

US-09-985-689a-3 (1-433) x AB046402 (1-1299)
Qy      1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnAsnAsnTyrGlyLeuTyr 20
Db      1 AATGATGTAGCAAGAGGAATAGTAAAGCAGACAGCGTTGCAACAAACAAATACCGACTATAT 60
Qy      21 GlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db      61 GGCACAGGTCAGTAGTTGGCAGTAGCGGATACGGGTTTAGATACAGGTCGTACCGATAGT 120
Qy      41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db      121 TCTATGATCAAGACATCCCGTGGGAAATATACAGCTCTTTACCGCTTAGGAAGAACTAAC 180
Qy      61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnAla 80
181 AATGCAAAATGATCCGATGGGATGGTACGATAGCTGGTCTGTGCTTGGTAAATGCT 240
81 LeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSerSer 100
241 TTAATAAAGGAATGGCTCCGAAGCTAACTTAGTCTTCCAATCTATTATTATGATGACG 300
101 GlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTyrAsnAla 120
301 GGGAGGATTAGGAGGATTACCATCGAATTAATACGTTATTATTAGTCAAGCTTGGATGCT 360
121 GlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrAlaAsn 140
361 GCGCTAGAAATTCATACTAACTCTTGGGGGGCCCAAGTAAATGGAGCGTACACTGCTAAC 420
141 SerArgGlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaGly 160
421 TCGAGACAGTGGATGATGATGTTTGAACACATGATATGACCGTACTTTTTCAGCTGGA 480
161 AsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThr 180
481 AATGAAGGACCTAACTCTGGAAACAATTAGCGCTCCAGGGACAGCGAAAAAATGCCATTAC 540
181 ValGlyAlaThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsnProAsnHis 200
541 GTCGGCGCACGGAAGAACTACCGACCAAGTTTGGTTCAATTGCAGTAAACCTTAATCAT 600
201 IleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspValThr 220
601 ATCGCACAAATTTTCATCGAGAGGAGCTACGAGAGATGGACGAATTAACACGACGTAACA 660
221 AlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTyr 240
661 GCTCCTGGAACTTTATATTATCAGCAGCGCTCTTCTTTAGCACCACTCTTCGTTTGG 720
241 AlaAsnTyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleVal 260
721 GCGAATTTATAACAGTAGTAGTATATGCGGGGCTACCTCTATGCGCACACCTATAGTT 780
261 AlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrProLys 280
781 GCGGGGAATTCGCGCAATTTACCGAGCATTTTATAAAAAATAGAGAAATTACACCTAAA 840
281 ProSerLeuIleLysAlaAlaLeuIleAlaGlyAlaThrAspValGlyLeuGlyTyrPro 300
841 CCTTCCTTATAAAGCTGCACTTATCGCTGGGGCTACTGATGTTGGTTAGGATATCCA 900
301 SerGlyAspGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyrVal 320
901 AGTGTGACCAAGCTGGGGCGGTGTACTTTAGATAAAATCGTTAAATGATGCGTATGTC 960
321 AsnGluAlaThrAlaLeuThrThrGlyGlnLysAlaThrTyrSerPheGlnThrGlnAla 340
961 AATGAAGCAACTGCATTAACAACAGGACAAAAGCAAGTATTCGTTCCAAACGCAAGCG 1020
341 GlyLysProLeuLysIleSerLeuValTrpThrAspAlaProGlySerThrThrAlaSer 360
1021 GGTAAACCATTAATAAATCTCGTTAGTATGACAGATGCACCTGGAAGTACACAGCATCT 1080
361 TyrThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyGlnLysTyrVal 380
1081 TATACACTAGTTAATGATTAGATCTAGTTATTACTGCTCCGAATGGACAAAAATATGTA 1140
381 GlyAsnAspPheSerTyrProTyrAspAsnAsnTrpAspGlyArgAsnValGluAsn 400
1141 GGTAAATGATTTAGTTATCTTATGATAATAATTTGGATGCTGCGCAACAATGTTGGAAC 1200
401 ValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsnVal 420
1201 GTATTTATAACGCTCCGCAATCTGGAACGATATACATTCAGGTTCAAGCGTATACGTT 1260
421 ProSerGlyProGlnArgPheSerLeuAlaIleValHis 433

```

```

Db      1261 CCATCTGACACACAGCGTTTCTCACTAGTATCGTACAT 1299

RESULT 2
LOCUS   AB046405                1299 bp DNA linear BCT 23-JAN-2001
DEFINITION Bacillus sp. SD521 PROD gene for protease, partial cds.
ACCESSION AB046405
VERSION   AB046405.1 GI:12381942
KEYWORDS .
SOURCE   Bacillus sp. SD521
ORGANISM Bacillus sp. SD521
REFERENCE 1 (sites)
AUTHORS  Saeki,K., Okuda,M., Hatada,Y., Kobayashi,T., Ito,S., Takami,H. and
          Hozikoshi,K.
TITLE    Novel oxidatively stable subtilisin-like serine proteases from
          alkaliphilic Bacillus spp.: enzymatic properties, sequences, and
          evolutionary relationships
JOURNAL  Biochem. Biophys. Res. Commun. 279 (2), 313-319 (2000)
MEDLINE  20568675
PUBMED   11118284
REFERENCE 2 (bases 1 to 1299)
AUTHORS  Saeki,K.
DIRECT SUBMISSION
SUBMITTED (20-JUL-2000) Katsuhisa Saeki, Kao corporation,
Biological Science Laboratory, Ichikaimachi Akabane 2606, Haga,
Tochigi 321-3497, Japan (E-mail:387185@kasanet.kao.co.jp,
Tel:81-285-68-7400, Fax:81-285-68-7403)
LOCATION/Qualifiers
1. .1299
   /organism="Bacillus sp. SD521"
   /mol_type="genomic DNA"
   /strain="SD521"
   /db_xref="taxon:133780"
1. .1299
   /gene="PROD"
   <1. .>1299
   /gene="PROD"
   /codon_start=1
   /transl_table=11
   /product="protease"
   /protein_id="BAB21268.1"
   /db_xref="GI:12381943"
   /db_xref="F04929"
   /translation="NDVARGIVKADVAQNNGLYGGQGVAVADTGLDGRNDSMHE
APRGIKITALYALGRNTNADPNHGHTVAGSVLGNALNKGWAPQANLVFOSIMDSGG
LGLPSNLNTLFSQAWNAGARIHNSWGPVNGAYTANSROVDYVNRNDYVLPAAG
NEGNSGIFLSPGAKXALITVQATENRPSFGSLADPNHIAQSFSGRATDRKRED
VIAPGTFILSAPSLAPDSFWANYNKSIATFMGTSMTPIVAGNVQALREHFINKRG
ITPKESLIKALIGATDVLGYPGSGQGRVTLDKSLNVAYVNEATALATQKATY
SFQAGAKPLKISLVITDAPGSTTASVTLVNDLDELITAPNGQRYVGNDFSPYDNNW
DGRNVENVFVNAQSGTYTIEVQAYNVPSGPQRFSLAIVH"
ORIGIN
Alignment Scores:
Pred. No.: 1,77e-133 Length: 1299
Score: 2251.00 Matches: 430
Percent Similarity: 99.54% Conservative: 1
Best Local Similarity: 99.31% Mismatches: 2
Query Match: 99.47% Indels: 0
DB: 1 Gaps: 0

US-09-985-689A-3 (1-433) x AB046405 (1-1299)

Qy      1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnAsnAsnTyrGlyLeuTyr 20
Db      1 AATGATGTAGCAAGAGGAATAGTAAAGACAGACGTTGCACAAAACAAATACGGACTATAT 60
Qy      21 GlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db      61 GGACAGGGTCAAGTAGTGTGACGTACGGATACGGTTTGTAGTACAGGTCGTAACGATAGT 120
Qy      41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyValArgThrAsn 60

```

| QY | 421 | ProSerGlyProGlnArgPheSerLeuAlaIleValHis | 433 |
|--|---|---|------------|
| QY | 421 | ProSerGlyProGlnArgPheSerLeuAlaIleValHis | 433 |
| DB | 1261 | CCATCTGGCCACAGCGTTTCTCATTAGTATCGTACAT | 1299 |
| RESULT 3 | | | |
| LOCUS | E03808 | 2218 bp | DNA linear |
| DEFINITION | E03808 | 2218 bp | DNA linear |
| ACCESSION | E03808 | 2218 bp | DNA linear |
| VERSION | E03808.1 | GI:2172022 | |
| KEYWORDS | JP 1992197182-A/1. | | |
| SOURCE | Bacillus sp. | | |
| ORGANISM | Bacillus sp. | | |
| REFERENCE | Bacteria: Firmicutes; Bacillales; Bacillaceae; Bacillus. | | |
| AUTHORS | 1 (bases 1 to 2218) | | |
| TITLE | Tobe, S., Odera, M., and Asai, Y. | | |
| JOURNAL | DNA CODING ALKALINE PROTEASE YA ENZYME AND PRODUCTION OF ALKALINE PROTEASE YA USING THE DNA | | |
| COMMENT | Patent: JP 1992197182-A 1 16-JUL-1992; LION CORP | | |
| OS | Bacillus sp. | | |
| PN | JP 1992197182-A/1 | | |
| PD | 16-JUL-1992 | | |
| PF | 28-NOV-1990 JP 1990327110 | | |
| PI | TOBE SEIICHI, ODERA MOTOYASU, ASAI YOSHIO | | |
| PC | C12N15/57, C12N15/386, C12N9/54, (C12N15/57, C12N9/54, C12N15/07), | | |
| PC | C12N15/07), | | |
| CC | strandedness: Double; | | |
| CC | topology: Linear; | | |
| CC | hypothetical: No; | | |
| CC | anti-sense: No; | | |
| CC | *source: strain=Y strain; | | |
| CC | *source: clone=pUB9A; | | |
| PH | Key | | |
| PH | Location/Qualifiers | | |
| FT | CDS | 218..2125 | |
| FT | /product='precursor of alkaline protease Ya' | | |
| FT | sig peptide | 218..823 | |
| FT | mat peptide | 824..2122 | |
| FT | /product='precursor of alkaline protease Ya' | | |
| FEATURES | | | |
| source | 1..2218 | | |
| ORIGIN | /organism="Bacillus sp." /mol_type="genomic DNA" /db_xref="taxon:1409" | | |
| Alignment Scores: | | | |
| Pred. No.: | 1..9e-132 | Length: | 2218 |
| Score: | 2239.00 | Matches: | 428 |
| Percent Similarity: | 99.31% | Conservative: | 2 |
| Best Local Similarity: | 98.85% | Mismatches: | 3 |
| Query Match: | 98.94% | Indels: | 0 |
| DB: | 6 | Gaps: | 0 |
| US-09-985-689A-3 (1-433) x E03808 (1-2218) | | | |
| QY | 1 | AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnAsnAsnTyrGlyLeuTyr | 20 |
| DB | 824 | AATGATGTACAGAGGATAGTAAAGCTGATGTGCACAAAACAATACGGATTATAT | 883 |
| QY | 21 | GlyGlnGlyValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer | 40 |
| DB | 884 | GGACAAAGGTCACACTAGTTGAGTACAGGACACAGGCTTAGATACAGGTCGTAACGATAGT | 943 |
| QY | 41 | SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn | 60 |
| DB | 944 | TCTATGCATGAACATTCGCGGAAATATCAAGCTCTTTACCGTTAGGAGAACTAAT | 1003 |
| QY | 61 | AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnAla | 80 |
| DB | 1004 | AATGCGAGTATCCGAATGGCATGGCAGACATGTAGCAGGTTCTGTACTTGTGTAATGCT | 1063 |

RESULT 4
AB046404
LOCUS AB046404 1299 bp DNA linear BCT 23-JAN-2001
DEFINITION Bacillus sp. Y PROC gene for protease, partial cds.
ACCESSION AB046404
VERSION AB046404.1 GI:12381940
KEYWORDS
SOURCE
ORGANISM Bacillus sp. Y
Bacillus sp. Y
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
1 (sites)
Saeki, K., Okuda, M., Hatada, Y., Kobayashi, T., Ito, S., Takami, H. and
Horikoshi, K.
TITLE Novel oxidatively stable subtilisin-like serine proteases from
alkaliphilic Bacillus spp.: enzymatic properties, sequences, and
evolutionary relationships
Biochem. Biophys. Res. Commun. 279 (2), 313-319 (2000)
JOURNAL
MEDLINE 20568675
PubMed 11118284
REFERENCE 2 (bases 1 to 1299)
AUTHORS Saeki, K.
DIRECT SUBMISSION
SUBMITTED (20-JUL-2000) Katsuhisa Saeki, Kao corporation,
Biological Science Laboratory, Ichikaimachi Akabane 2606, Haga,
Tochigi 321-3497, Japan (E-mail:387185@kasanet.kao.co.jp,
Tel:81-285-68-7400, Fax:81-285-68-7403)
FEATURES
source
1..1299
/organism="Bacillus sp. Y"
/mol_type="genomic DNA"
/strain="Y"
/db_xref="taxon:133779"
1..1299
/gene="PROC"
<1..>1299
/gene="PROC"
/codon_start=1
/evidence="experimental"
/transl_table=1
/product="protease"
/protein_id="BAB21267.1"
/db_xref="GI:12381941"
translation="NDVARGIVKADVAQNNGYVGQGVAVADTGLDTRNDSSMHE
AFRGKITALYALGNTNADPNHGTHVAGSVLGNALNGKMAPQANLVFOSIMDSGG
LGLPLNLTLPFOANWAGARIHTNSWGPVNGATNSROVDYVNRNDMTVLFAAG
NEGPNSTGISAPGTAKNAITVGATENVRPSPGSTADPNHIAOFSSFGATEDRIKPD
VTAGTFTLSRSLAPDSFWANYNSKYAMGTSMTATPIVAGNVQLREHETIKRG
ITPRFLKALIIAGIDVGLGYPNGDQGWGRVTLNKSINVAIVNEATALATGKATY
SFOAGAKPLKISLVMTDAPGSTTASTYTLVNDLVLITAPNGQKYGVNDFSPYDNNW
DGRNENVFINAPQSGTYIIIEVOAYNVPSPQRFSLAIVH"

gene

CDS

Alignment Scores:
Pred. No.: 2, 1e-132 Length: 1299
Score: 2234.00 Matches: 427
Percent Similarity: 99.31% Conservative: 3
Best Local Similarity: 98.61% Mismatches: 3
Query Match: 98.72% Indels: 0
DB: 1 Gaps: 0
US-09-985-689A-3 (1-433) x AB046404 (1-1299)

ORIGIN

1 AenAspValAlaArgGlyLeuValValAlaAspValAlaGlnAenAsnTyrGlyLeuTyr 20
Db 1 AATGATGATGACGAAGGATAGTAAAGCTGATGTTGCACAAAACAATTACGGATTATAT 60
Qy 21 GlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 51 GCACCAAGGTCAAGTAGTTCGAGTAGCGACACAGGCTTAGATACAGGTCGTAACGATAGT 120
Qy 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 121 TCTATGATGATGAGCATTCGCGGGGAAAATCACAGCTCTTTACGCGTTAGGAAGAACTAAT 180

61 AenAlaAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAenAla 80
Db 181 AATGCCAGTATCCGAATGGCATGGCACACATGTAGCAGGTTCTGTATTGTAATGCT 240
Qy 81 LeuAsnLysGlyMetAlaProGlnAlaAenLeuValPheGlnSerIleMetAspSerSer 100
Db 241 TTAATAAAGGAATGGCTCCGCAAGTAACCTTAGTCTTCCAAATCTATTATGATAGCAGC 300
Qy 101 GlyGlyLeuGlyGlyLeuProSerAenLeuAenThrLeuPheSerGlnAlaTrpAenAla 120
Db 301 GGAAGGATTAGTGCTTACCATCGAACTTAATAATACGTTATTATTAGTCAAGCTTGGAAATGCT 360
Qy 121 GlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyValaTyrThrAlaAsn 140
Db 361 GGAGCAAGAATTCATCTAATCTTGGGAGCCCAAGTAATGAGCGGTACACTGCTAAC 420
Qy 141 SerArgGlnValAspGluTyrValArgAenAsnAspMetThrValLeuPheAlaAlaGly 160
Db 421 TCAGACAAGTGGATGAATATGTTCCAAATAATGATATGACGGTACTTTTTTCAGCTGCT 480
Qy 161 AsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThr 180
Db 481 AATGAAGGCTCTAATTCAGGAACAATTAGTGTCCAGGTACAGCGAAAAAATGCTATTACG 540
Qy 181 ValGlyAlaThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsnProAsnHis 200
Db 541 GTCGGCGCAACGAAAACTATCGCCCAAGCTTCGGTTCGATAGCAGATAACCCAAATCAT 600
Qy 201 IleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyValGlyLeuLysProAspValThr 220
Db 601 ATTGCACAATTTTCATCGAGAGGAGCTACGAGGATGGACGAATTAAGCTCAGCTAAC 660
Qy 221 AlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrp 240
Db 661 GCTCTCTGGAACATTTATTTTATCAGCACGTTCTTCTCTAGCTCCAGACTCTTCGTTTGG 720
Qy 241 AlaAsnTyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleVal 260
Db 721 CGAATTATTAACAGTAATACCGGTATATGGCGGTACCTCCATGCGGCACACTATTGTT 780
Qy 261 AlaGlyAsnValaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrProLys 280
Db 781 GCAGGGAATGTCGCGCAATTAGCTGAGCATTTTATAAAAAATAGAGGTATTACTCTCTAAG 840
Qy 281 ProSerLeuIleLysAlaAlaLeuIleAlaGlyAlaThrAspValGlyLeuGlyTyrPro 300
Db 841 CCTTCTTTAATAAAGCTGCACTTATCGCTGTGCTACTGATGTTGGTTAGGATATCT 900
Qy 301 SerGlyAspGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValaTyrVal 320
Db 901 AATGGTGACCAAGGCTGGGGCGGTGTTACTCTAAATAAATCGTTAAATGTAGCGTATGTC 960
Qy 321 AsnGluAlaThrAlaLeuThrThrGlyGlnLysAlaThrTyrSerPheGlnThrGlnAla 340
Db 961 AATGAAGCAACTGCATTAGCCACAGACAAAAGCAACGATATTCGTTCACACCAAGCG 1020
Qy 341 GlyLysProLeuLysIleSerLeuValTrpThrAspAlaProGlySerThrThrAlaSer 360
Db 1021 GGTAAACCTTTAAAAATCTCGTTAGTAGACAGATGCTCCTCGGAAGTACAACTGCATCT 1080
Qy 361 TyrThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyGlnLysTyrVal 380
Db 1081 TATACACTAGTTAATGAATTAGATCTAGTTATTACTGCTCCGAATGGCAAAAATATGTA 1140
Qy 381 GlyAsnAspPheSerTyrProTyrAsnAenAsnTrpAspGlyArgAsnAenValGluAsn 400
Db 1141 GGAATAGATTTTAGTTATCTTATGATAATACTGGGATGGTCGCAACAATGTTGAGAAC 1200
Qy 401 ValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsnVal 420
Db 1201 GTATTTTAAACGCTCCGCAATCTGGAACGTATATAATTAGGTTCAAGCGGTATAATGTA 1260

```
QY 421 ProSerGlyProGlnArgPheSerLeuAlaIleValHis 433
DB 1261 CCATCTGGCCCAAGCGTTTCTCACTAGCTATCGTACAT 1299

RESULT 5
LOCUS AR069954 3003 bp DNA linear PAT 18-FEB-2000
DEFINITION Sequence 41 from patent US 5891701.
ACCESSION AR069954
VERSION AR069954.1 GI:7220842
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE
AUTHORS Sloma,A. and Christianson,L.
TITLE Nucleic acids encoding a polypeptide having protease activity
JOURNAL Patent: US 5891701-A 41 06-APR-1999;
FEATURES
LOCATION/Qualifiers
1..3003
/morganism="unknown"
/mol_type="unassigned DNA"
ORIGIN
Alignment Scores:
Pred. No.: 1e-119 Length: 3003
Score: 2040.00 Matches: 385
Percent Similarity: 94.46% Conservative: 24
Best Local Similarity: 88.91% Mismatches: 24
Query Match: 90.15% Indels: 0
DB: 6 Gaps: 0

US-09-985-689A-3 (1-433) x AR069954 (1-3003)
QY 1 AsnAspValAlaArgGlyIleValHisAlaAspValAlaGlnAsnTyrGlyLeuTyr 20
DB 1470 AATGACGTGGCCGTTGGCATTGTGAAGCAGACGTCGCACAAATAACTTTGGCTTATAT 1529
QY 21 GlyGlnGlyValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
DB 1530 GGACAGGACAGATTGTAGCAGTTGCTGATCTGGCTTGATACAGAGAAATGACAGT 1589
QY 41 SerMetHisGlnAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
DB 1590 TCGATGCATGACGATTCGCGGTAGATTACCGCTATATGCTATGCTGGCAGACGAT 1649
QY 61 AsnAlaAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnAla 80
DB 1650 AACGCCAATGATCCAAATGGACATGGAAACCCATGTTGCTGGATCTGTGTAGAAATGCT 1709
QY 81 LeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSerSer 100
DB 1710 ACAATTAAGGATGGACCGCAGCAACATCTAGTCTTCAATCTATTATGATAGTGT 1769
QY 101 GlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsnAla 120
DB 1770 GGAGGGCTGGAGGACACCTGCTGATCTACAAACATTTATTCAGTCAAGCATATAGTGT 1829
QY 121 GlyAlaArgIleHisThrAsnSerTrpGlyValProValAsnGlyValAlaTrpAlaAsn 140
DB 1830 GGACGAGATTCATACGATTTATGCGGGGCTCCAGTAAACGGTCCCTATACGACAGAC 1889
QY 141 SerArgGlnValAspGluTyrValArgAsnAspMetThrValLeuPheAlaAlaGly 160
DB 1890 TCTCGAAATGTTGATGATTATGTGAGAAAAAATGATATGACGATCTTTTGGCGCGGA 1949
QY 161 AsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThr 180
DB 1950 AATGAGGACAGTACGCGGTACCAATCAGTGCACCGAGAAACAGCAAAATGCGATTACA 2009
QY 181 ValGlyAlaThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsnProAsnHis 200
DB 2010 GTTGGGCAACCGAAACCTACGTCCTCGGATCTTATCGGATATATTAACCAT 2069
```

```
QY 201 IleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyValGlyLeuLysProAspValThr 220
DB 2070 GTTGTCTCAATTTCTTTCACAGGTCCTACTAGATGGACGTATTAAAGCGGACGTCATG 2129
QY 221 AlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrp 240
DB 2130 GCACCAAGGTACGTATATTCTCTCTGTAGATCATCATTAAGTCCAGATTCTCTATCTCG 2189
QY 241 AlaAsnTyrAsnSerLysIleValAlaTyrMetGlyThrSerMetAlaThrProIleVal 260
DB 2190 GCAACCATGATAGTAATAATGCTCATCGGTGGTACTTCTTATGGCTACTCCATTGTA 2249
QY 261 AlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrProLys 280
DB 2250 GCAGGTAACTTGCACAAATTAAGGAGCATTTTGTGAAAATAGAGGGGTAACTCTCTAAG 2309
QY 281 ProSerLeuIleLysAlaAlaIleAlaGlyAlaThrAspValGlyLeuGlyTyrPro 300
DB 2310 CCTTCCCTTTTAAAGCTGCTTTAATGTCAGGTGCTGCGGATGTTGGACTTGGCTTTCCA 2369
QY 301 SerGlyAspGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaLysVal 320
DB 2370 AATGGTAACCAAGGATGGGAGAGTAACGTTAGATAAATCCCTAAATGTGCGATTGTG 2429
QY 321 AsnGluAlaThrAlaLeuThrThrGlyGlnLysAlaThrTyrSerPheGlnThrGlnAla 340
DB 2430 AATGAAACGAGCCCTTTATCAACAGTCAAAAGCAACATATTCGTTTACGGCTCAAGCT 2489
QY 341 GlyLysProLeuLysIleSerLeuValTrpThrAspAlaProGlySerThrThrAlaSer 360
DB 2490 GGTAACCCCTTAAATATATCACTTGTGTCAGATGCACAGGTAGCAGCGGCATCA 2549
QY 361 TyrThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyGlnLysTyrVal 380
DB 2550 CTAACCTTTAGTGAATGATTAGACTTAGTAATCACTGCACCAATGGAACCTAAATACGTC 2609
QY 381 GlyAsnAspPheSerTyrProTyrAspAsnAsnTrpAspGlyArgAsnAsnValGluAsn 400
DB 2610 GGAATGACCTTACAGCACCGGTATGATAACAATGGATGGCAGAACACACGTTGGAAT 2669
QY 401 ValPheLeuAsnAlaProGlnSerGlyThrTyrIleGluValGlnAlaLysVal 420
DB 2670 GTGTTTATCAATGCTCTCTAAAGCGGACGTATACAGTCGAAGTCGAGGCTTCAATGTA 2729
QY 421 ProSerGlyProGlnArgPheSerLeuAlaIleValHis 433
DB 2730 CCAGTAAGTCGCAACCTTTCTTTAGCGATTGTACAT 2768

RESULT 6
BD062155 3003 bp DNA linear PAT 27-AUG-2002
LOCUS Nucleic acids encoding a polypeptide having protease activity.
DEFINITION
ACCESSION BD062155
VERSION BD062155.1 GI:22607760
KEYWORDS JP 2001514529-A/39.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE
AUTHORS Sloma,A. and Christianson,L.
TITLE Nucleic acids encoding a polypeptide having protease activity
JOURNAL Patent: JP 2001514529-A 39 11-SEP-2001;
COMMENT
NOVO NORDISK BIOTECH INC
PN JP 2001514529-A/39
PD 11-SEP-2001
PF 09-JUN-1998 JP 1999503145
PR 12-JUN-1997 US 08/873479
PI ALAN SLOMA,LYNNE CHRISTIANSON
PC C12N15/57,C12N15/75,C12N9/54,C12K14/00
CC Strandedness: Single;
CC Topology: Linear;
```

FEATURES

source

Location/Qualifiers

1..3003

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

ORIGIN

Alignment Scores:

Pred. No.: 1e-119 Length: 3003
Score: 2040.00 Matches: 385
Percent Similarity: 94.46% Conservative: 24
Best Local Similarity: 88.91% Mismatches: 24
Query Match: 90.15% Indels: 0
DB: 6 Gaps: 0

US-09-985-689A-3 (1-433) x BD062155 (1-3003)

QY 1 AsnAspValAlaArgGlyLeuValValAlaAspValAlaGlnAsnAsnTyrGlyLeuTyr 20
DB 1470 AATGACGTGGCGCGTGGCAATGTGAAGCAGACGCTCGCACAAATAAATTTGGCTTATAT 1529
QY 21 GlyGlnGlyGlnValValAlaAlaAspThrGlyLeuAspThrGlyArgAspSer 40
DB 1530 GCACAGGACAGATTTGTAGCGTGTCTGATCTGGGCTTGATACAGGAGAAATGACAGT 1589
QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
DB 1590 TCGATGATGAAGCATTCGCGGTGAAGATTACCGCACTATATGCACTGGGCGAGAACGAAT 1649
QY 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnAla 80
DB 1650 AACGCCAATGATCCAAATGGACATGGAAACCAATGTTGCTGATCTGTGTTAGGAAATGCT 1709
QY 81 LeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSerSer 100
DB 1710 ACAATAAAGGATGGCAGCGCAACCAATCTAGTCTTTTCAATCTATTATGATAGTGGT 1769
QY 101 GlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTyrPheAla 120
DB 1770 GGAGGGCTGGGAGGACTGCTCTAATCTACAAACATTTATTCAGTCAAGCATATAGTGGT 1829
QY 121 GlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrAlaAsn 140
DB 1830 GGAGCGAGATTCATACGATTCATGGGGGCTCCAGTAACGGTGCCTATACGACAGAC 1889
QY 141 SerArgGlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAlaGly 160
DB 1890 TCTCGAAATGTTGATGATATGTGAGAAAAAATGATGACGATTTCTTTTTCGGCGCGGA 1949
QY 161 AsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThr 180
DB 1950 AATGAGGACCGAGTAGCGGTACATCATGTGACCGAGAACAGCAAAAATGCGATTACA 2009
QY 181 ValGlyAlaThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsnProAsnHis 200
DB 2010 GTTGGGCAACCGAAACCTAGCTCCAGCTTCGGATCTTATGCGGATTAATATTAAACCAT 2069
QY 201 IleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgGlyLeuLysProAspValThr 220
DB 2070 GTTGCTCAATTTCTTCAGAGGCTCTACTAGATGAGCAGGATTAAACCGAGCTGATG 2129
QY 221 AlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrp 240
DB 2130 GCACCAAGTACGTATATTCTCTGCTAGATCATCATTAGCTCCAGATTCCTCATTTCTGG 2189
QY 241 AlaAsnTyrAsnSerLysIleValAlaTyrMetGlyGlyThrSerMetAlaThrProIleVal 260
DB 2190 GCAACCATGATAGTAAATATGCTACATGAGGGTGTACTTCTATGGCTACTCCAAATTGTA 2249
QY 261 AlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrProLys 280
DB 2250 GCAGGTAAATGTCACAAATTAGGGAGCATTTGTGAAAAATAGAGGGGTAACTCTCTAAG 2309

QY 281 ProSerLeuIleLysAlaAlaLeuIleAlaGlyAlaThrAspValGlyLeuGlyTyrPro 300
DB 2310 CTTTCCCTTTTAAAGCTGCTTTTAAATTCAGGTGTCGGGATGTTGGACTTGGCTTTTCA 2369
QY 301 SerGlyAspGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyrVal 320
DB 2370 AATGGTAACCAAGGATGGGAAGAGTAACGTTAGATAAAATCCCTAAATGTCGCAATTTGTG 2429
QY 321 AsnGluAlaThrAlaLeuThrThrGlyGlnLysAlaThrTyrSerPheGlnThrGlnAla 340
DB 2430 AATGAACCGAGCCCTTTTCAACAAGTCAAAAGCAACATATTCGTTTACGGCTCAAGCT 2489
QY 341 GlyLysProLeuLysIleSerLeuValThrThrAspAlaProGlySerThrThrAlaSer 360
DB 2490 GGTAAACCCCTTAAATAATATCACTTGTGTCAGATGCACAGGTAGCAGCAGCGCATCA 2549
QY 361 TyrThrLeuValAsnAspLeuValIleThrAlaProAsnGlyGlnLysTyrVal 380
DB 2550 CTAACCTTTAGTGAATGATTAGACTTAGTAATCACTGCACCAATGGAACTAAATACGTC 2609
QY 381 GlyAsnAspPheSerTyrProTyrAspAsnAsnTrpAspGlyArgAsnAsnValGluAsn 400
DB 2610 GGAAATGACTTTACAGCACCGTATGATAACAATTTGGATGGCAGAAACAACCTGGAAAT 2669
QY 401 ValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsnVal 420
DB 2670 GTGTTTATCAATGCTCTCTCAAGCGGACGTATACAGTCGAAGTGCAGGCTTACAATGTA 2729
QY 421 ProSerGlyProGlnArgPheSerLeuAlaIleValHis 433
DB 2730 CCAGTAAGTCCGAAACCTTTCTTTAGCGATTGTACAT 2768
RESULT 7
AB046406 1302 bp DNA linear BCT 23-JAN-2001
LOCUS Bacillus sp. NV1 PROE gene for protease, partial cds.
DEFINITION
ACCESSION AB046406
VERSION AB046406.1 GI:12381944
KEYWORDS
SOURCE Bacillus sp. NV1
ORGANISM Bacillus sp. NV1
REFERENCE 1 (sites)
AUTHORS Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
Horikoshi, K.
Saeki, K., Okuda, M., Hatada, Y., Kobayashi, T., Ito, S., Takami, H. and
Novel oxidatively stable subtilisin-like serine proteases from
alkaliphilic Bacillus spp.: enzymatic properties, sequences, and
evolutionary relationships
Biochem. Biophys. Res. Commun. 279 (2), 313-319 (2000)
JOURNAL
MEDLINE 20568675
PUBMED 11118284
REFERENCE 2 (bases 1 to 1302)
AUTHORS Saeki, K.
TITLE Direct Submission
SUBMITTED (20-JUL-2000) Katsuhisa Saeki, Kao corporation,
Biological Science Laboratory, Ichikaimachi Akabane 2606, Haga,
Tochigi 321-3497, Japan (E-mail:387185@kastanet.kao.co.jp,
Tel:81-285-68-7400, Fax:81-285-68-7403)
FEATURES
source
1..1302
/organism="Bacillus sp. NV1"
/mol_type="genomic DNA"
/strain="NV1"
/db_xref="taxon:133781"
1..1302
/gene="PROE"
1..1302
/gene="PROE"
/codon_start=1
/transl_table=11
/product="protease"
/protein_id="BAB21269.1"


```

/db_xref="GI:12381945"
/translation="NDVARGIVKADVAQSSYGLYGGQGVAVADTGLDTRNDSSMHE
AFRGKITAVALGRNNDNGHCHTVAGSVLNGTSGKMGAPQANLVFQSVYDNG
GLGHPNSVTLFQSAVAGAR IHTNSGAPVNGAYTTDSRVDVYRKNDMVLPAA
GNEGNGTISAPGTAKNAITVGENLRPSFGYADNIHVAQFSRGPTKGRIPK
DVMAPGTILSARSLADPSFVWHDYSKAYMGTSNATPIVAGNVQALREFIKPR
GITPKSLKALAGATDILGUGPSGQNGRVTLDKSNLVAFNVNTESSLTNQKAT
YSFTAQSGKPLKISLWSDAPASTSASVTLVNDLDTLVPNGTKYVGNDFTPAYDNN
WDGRNENVFNAPQSGTYTVEQAVNVPOQPOAFSLAIVN"

```

ORIGIN

```

Alignment Scores:
Pred. No.: 6,366-119 Length: 1302
Score: 2020.50 Matches: 383
Percent Similarity: 94.70% Conservative: 28
Best Local Similarity: 88.25% Mismatches: 22
Query Match: 89.28% Indels: 1
DB: 1 Gaps: 1

US-09-985-689A-3 (1-433) x AB046406 (1-1302)

Qy 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnAsnAsnTyrGlyLeuTyr 20
Db 1 AACGATGTTGCCAGAGCATTGTAAAAGCCGATGTTGCCAGAGCAGCTATGGTTTATAT 60
Qy 21 GlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyValArgAsnAspSer 40
Db 61 GGCACAGGCAAGTGTTCAGTAGCGGATACCGACTGGATACAGGCGGTAAATGACAGC 120
Qy 41 SerMetHisGluAlaPheArgGlyIleThrAlaLeuTyrAlaLeuGlyValArgThrAsn 60
Db 121 TCGATGCATGAGCGTTCCGAGGAAAGATTACCGCGATATATGCCCTTGGAGAACAAAC 180
Qy 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyValAsnAla 80
Db 181 AACGCCAATGATCCAAACCGACACGGACGCATGTTGCCGGATCTGTTTAGAAACCGT 240
Qy 81 Leu---AsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 99
Db 241 ACTTCGATAAAGGATGCTCCACAAAGCTAATCTAGTTTTCATCTGTTATGACAGC 300
Qy 100 SerGlyGlyLeuGlyLeuProSerAsnLeuAsnThrIleuPheSerGlnAlaTrpAsn 119
Db 301 AATGGCGGACTTCGGCGACTTCCTTCCAAATGTATGATACATTATTCAGCCAGCATATAGT 360
Qy 120 AlaGlyValaArgIleHisThrAsnSerTrpGlyValaProValAsnGlyAlaTyrThrAla 139
Db 361 GCCGGTCCAGATCCATACGAACTCATGGGAGCGCTGTAAACGAGCGTACACTACT 420
Qy 140 AsnSerArgGlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAla 159
Db 421 GATTCCAGAAACGTAGACGATTATGTTCTGTAATAATGATATGCGCGTCTCTTTTGCAGCG 480
Qy 160 GlyAsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 179
Db 481 GGTAAACAGAGGCGGAATGCGGACACATCAGTGTCTCTGTAACCGGAGATGCTATC 540
Qy 180 ThrValGlyAlaThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsnProAsn 199
Db 541 ACAGTAGGGGCAACAGAAAACCTGCGCCCAAGCTTTGGATCTTATGCTGACACATCAAT 600
Qy 200 HisIleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyValArgIleLysProAspVal 219
Db 601 CATGTACACAGTTTCTCTCCGCGGACCTTACAAAGATGGATGATATCAACCGGACGTA 660
Qy 220 ThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 239
Db 661 ATGGCACCCAGGAACATTATTTATCGCAAGATCTTCTTGGCTCCGACTCCTCATTC 720
Qy 240 TrpAlaAsnTyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 259
Db 721 TGGGCAACCAATGACGAAATATGCTTATATGGGTGGAAACATCCATCGGCGACACCGATT 780

```

```

Qy 260 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro 279
Db 781 GTAGCTGGTAACGTTGCCAGCTAGCTGAACATTTCAACAAAAACAGAGGAATCACTCCT 840
Qy 280 LysProSerLeuIleLysAlaAlaLeuIleAlaGlyAlaThrAspValGlyLeuGlyTyr 299
Db 841 AAACCAATCTTGTGTAAGCAGCTCTTATGCGGAGCAACTGATATCGTCTTGGCTAT 900
Qy 300 ProSerGlyAspGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 319
Db 901 CCGAGTGGAAACCAAGGATGGGAGAGTAACATGGCAAGCTCACTTAATGACTTTC 960
Qy 320 ValAsnGluAlaThrAlaLeuThrThrGlyGlnLysAlaThrTyrSerPheGlnThrGln 339
Db 961 GTAAATGAACAAGCTCTTTACTACTACTACCAAAAGGCTACGTATTCATTTACTGCACA 1020
Qy 340 AlaGlyLysProLeuLysIleSerLeuValTrpThrAspAlaProGlySerThrThrAla 359
Db 1021 TCAGGCACAACTTTGAAGATTTTCATTTGGTTTGGTCTGTATGTCACCGCAAGTACTTCCGCA 1080
Qy 360 SerTyrThrLeuValAsnAspLeuValIleThrAlaProAsnGlyGlnLysTyr 379
Db 1081 TCGGTTACATGTTGTAATGATCTGGATCTGGTGATTCAGCTCCCAATGGACAAAGTAT 1140
Qy 380 ValGlyAsnAspPheSerTyrProTyrAspAsnAsnTrpAspGlyArgAsnAsnValGlu 399
Db 1141 GTTGGAAACGACTTACTCTCTCCCTATGATAATAACTGGGATGACGTAACATGTAGAG 1200
Qy 400 AsnValPheIleAsnAlaProGlnSerGlyThrTyrIleGluValGlnAlaTyrAsn 419
Db 1201 AAGTGTTTATCAATGCTCCGAAAGCGAAGCTATACAGTTGAGGTACAGGCTTACAT 1260
Qy 420 ValProSerGlyProGlnArgPheSerLeuAlaIleValHis 433
Db 1261 GTTCCACAGCGCGGAGCGGTTTTCTTTGGCTATTGTGAAC 1302

RESULT 8
AB084155 1923 bp DNA linear BCT 02-SBP-2003
LOCUS Bacillus sp. KSM-9865 gene for protease, complete cds.
DEFINITION
AB084155
ACCESSION
AB084155.1 GI:34392386
VERSION
AB084155.1
KEYWORDS
Bacillus sp. KSM-9865
Bacillus sp. KSM-9865
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
ORGANISM
1
REFERENCE
1 Okuda,M., Saeki,K. and Kobayashi,T.
AUTHORS
Bacillus sp. KSM-9865 protease gene
TITLE
Published Only in Database (2003)
JOURNAL
2 (bases 1 to 1923)
REFERENCE
Okuda,M., Saeki,K. and Kobayashi,T.
AUTHORS
Direct Submission
TITLE
Submitted (19-APR-2002) Mitsuyoshi Okuda, Kao corporation,
JOURNAL
Biological Science Laboratory; Ichikaimachi Akabane 2606, Haga,
Tochigi 321-3497, Japan E-mail:okuda.mitsuyoshi@kco.co.jp,
Tel:81-285-68-7543, Fax:81-285-68-7547
FEATURES
Location/Qualifiers
1..1923
/organism="Bacillus sp. KSM-9865"
/mol_type="genomic DNA"
/strain="KSM-9865"
/db_xref="taxon:192495"
1..1923
/codon_start=1
/transl_table=11
/product="protease"
/protein_id="BAC82522.1"
/db_xref="GI:34392387"
/translation="MRKKKVFSLVLSAAAII LSTVLSNPSAGARNFIDFKIQIT
TDAGFKQGGTGAALFVSENVKPKGLQKLETPANNKLIHVQNGPILEETKQ
OLEKTAKILDYIPDYAVIVEYEGDVKSATSTIEDVESVEPYLPFIYDIPQFTKAS
ELVAVALDITNQKNKEVQLRGIEIAQFATSNVDLYITAKPEYKVMNDVARGIVKADV

```

AQSSVGLYVQGGQIVAVADTGLDTRGNDSSMHEAPRGKITALYALGRTRNNANDTNGHGT
HVAGSVLNGSTNGKMAQANLVQSIIMDSGGGLGGLPSNLQTLFSAOYAGARIIHTN
SLRPAFVAGVATDSDNRVDYVRKNDMTILFPAAGNEPNGGTISAPGTAKNAITVGAIE
NKPFGGVDYADININVAQFSGRPTKDGRIKPDVMAQPTFILSARSLAPDSFPWANH
DSKYAIMGSTSMATPIVAGNVAQREHFVKRGITPKPSLLKALJAGAADIIGLYEN
GNQGVYTLDPKSLNVAYNSSSLSTSQKATYSFTACGKPLKLSIWSDDAPASTTA
SVTLVNDLULVITAPNGTQYVGNDFTSFINNWDGRNVNFINAQSQSYTILEVOA
INVPVGPOTFSLAIVN"

ORIGIN

Alignment Scores:

Pred. No.: 2,13e-117 Length: 1923
Score: 1999.50 Matches: 381
Percent Similarity: 94.01% Conservative: 27
Best Local Similarity: 87.73% Mismatches: 25
Query Match: 88.36% Indels: 1
DB: 1 Gaps: 1

US-09-985-689A-3 (1-433) x AB084155 (1-1923)

Qy 1 AsnAspValAlaArgGlyValLeuValAlaAspValAlaGlnAsnAsnTyrGlyLeuTyr 20
Db 619 AATGATGTTCAGCGTGAATGTCAAAGCGATGTGGCGCAGCAGCAGCTACGGGTGTAT 678
Qy 21 GlyGlnGlyGlnValValAlaAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 679 GGCAGGACACATCGTAGCGGTCCCGATACAGGGCTTGATACAGGTTCGAATGACAGT 738
Qy 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 739 TCGATGTCATGAAGCCCTCCGGGGGAAATATCTGCATTATATGCAATGTCGAGCGACCAAT 798
Qy 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn--- 79
Db 799 AATGCCATGATACGATGGTCATGTCAGCATGTGGCTCGCTCGATTAGGAACCGC 858
Qy 80 AlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 99
Db 859 TCCACTAATAAAGGAATCGCGCTCAGGCGAATCTAGTCTTCCCAATCTATCATGATGATG 918
Qy 100 SerGlyGlyLeuGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTyrAsn 119
Db 919 GGTGGGGNCTTGAGGACTACCTTCGAATCTGCAACCTTATTTCAGCAAGCATACAGT 978
Qy 120 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyValAlaTyrAla 139
Db 979 GCTGTGCGCAGAAATTCATCAAACTCTCGGGAGCAGCAGTGAATGGGCTTACACAACA 1038
Qy 140 AsnSerArgGlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAla 159
Db 1039 GATTCAGAAATGTGGATGATGATGTCGCAAAATGATAGACGATCCCTTTTCGCTGCC 1098
Qy 160 GlyAsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 179
Db 1099 GCGAATGAAGACCGAAGCGGGAACCATCATGTCAGCCAGCAGCAGCTAAATAATGCAATA 1158
Qy 180 ThrValGlyAlaThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsnProAsn 199
Db 1159 ACAGTCGGAGCTACCGAAACCTCCGCCCAAGCTTCGGGTCTTATGCGCAAAATATCAAC 1218
Qy 200 HisIleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspVal 219
Db 1219 CATGTGGCAGCATCTCTTCAGTGGCCGACAGAGATGGACGGATCAAAACCGGATGTC 1278
Qy 220 ThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 239
Db 1279 ATGGCAGCGGGAACGTTCTATATATCAGCAAGATCTCTCTTGGCAGCGATTCCTCTTC 1338
Qy 240 TrpAlaAsnTyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 259
Db 1339 TGGGCGAACCATGACAGTAAATATGATATCATGCGTGGAACTGCTCATGCTACACCGATC 1398
Qy 260 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro 279

Db 1399 GTTGTCTGGAACGTGCGACAGCTTCGTGACGATTTTGTGAAAACAGAGCGCATCACCA 1458
Qy 280 LysProSerLeuIleLysAlaAlaLeuIleAlaGlyValThrAspValGlyLeuGlyTyr 299
Db 1459 AAGCTTCTCTATTAAAGCGGCACTGATTCGGGTGCGAGCAGACATCGGCTTGGCTAC 1518
Qy 300 ProSerGlyAspGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 319
Db 1519 CCGACCGTAACCAAGGATGGGACGAGTGACATTGGATAAATCCCTAAACGTTGCCTAT 1578
Qy 320 ValAsnGluAlaThrAlaLeuThrGlyGlnLysAlaThrTyrSerPheGlnThrGln 339
Db 1579 GTGACGAGTCCAGTCTCTATCCACCACCCAAAGACGACGTACTCTGTTTACTGCTACT 1638
Qy 340 AlaGlyLysProLeuLysIleSerLeuValTrpThrAspAlaProGlySerThrThrAla 359
Db 1639 GCGGCAAGCCTTTGAAGATCTCCCTGGTATGCTGTGATGCCCTCGCAGCAACTGCT 1698
Qy 360 SerTyrThrLeuValAsnAspLeuValIleThrAlaProAsnGlyGlnLysTyr 379
Db 1699 TCCGTAAACGCTTGTCAATGACCTGGACCTTGTCATTACCGCTCCAAATGGCACACATAT 1758
Qy 380 ValGlyAsnAspPheSerTyrProTyrAspAsnAsnTrpAspGlyArgAsnAsnValGlu 399
Db 1759 GTTGGAAATGACTTTACTTCCGCATACATAATACTGGATGGCGCAATTAACGTAGAA 1818
Qy 400 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 419
Db 1819 AATGTATTATTAAATCGCCACAAAGCGGAGCGTATACAAATGAGGTACAGGCTTTAAAC 1878
Qy 420 ValProSerGlyProGlnArgPheSerLeuAlaIleValHis 433
Db 1879 GTACCGTTCGACACAGACCTTCTCGTTGGCAATTGTGAAC 1920
RESULT 9
AB046403
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source
gene
CDS
/organism="Bacillus sp. 9860"
/mol_type="genomic DNA"
/strain="9860"
/db_xref="taxon:133778"
1. 1920
/gene="PROA"
1. 1920
/gene="PROA"
/codon_start=1
AB046403
Bacillus sp. 9860
Bacillus sp. 9860
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
1
Saeki, K., Okuda, M., Hatada, Y., Kobayashi, T., Ito, S., Takami, H. and
Horikoshi, K.
Novel oxidatively stable subtilisin-like serine proteases from
alkaliphilic Bacillus spp.: enzymatic properties, sequences, and
evolutionary relationships
Biochem. Biophys. Res. Commun. 279 (2), 313-319 (2000)
20568675
11118284
2 (bases 1 to 1920)
Saeki, K.
Direct Submission
Submitted (20-JUL-2000) Katsuhisa Saeki, Kao corporation,
Biological Science Laboratory, Ichikimachi Akabane 2606, Haga,
Tochigi 321-3497, Japan (E-mail:387185@kasanet.kao.co.jp,
Tel:81-285-68-7400, Fax:81-285-68-7403)
On May 9, 2002 this sequence version replaced gi:12381938.
Location/Qualifiers
1. 1920
/organism="Bacillus sp. 9860"
/mol_type="genomic DNA"
/strain="9860"
/db_xref="taxon:133778"
1. 1920
/gene="PROA"
1. 1920
/gene="PROA"
/codon_start=1

```
/transl_table=11
/product="protease"
/protein_id="BAB21266.2"
/db_xref="GI:20521153"
/translation="WRKKVFLSVLSAAALSTVALNPNPSAGDARTFDLDFKGIQTTF
DVSFGSKQRTGAAFLVSESNVLLKLLKLETPANNKLHIVQFNGLIBETKQK
LTKGALIDVYIPDYAVIVVEGVDQSVRSIEHVESVEPYLPKYKIDPDQFTKGAST
LYKALALDPKONKSVQURGIIEIAYVASNDVHYITAKPEYKVMNDVARGIVKADVA
QSYGLYGGQIVAVADTGLDGRNDSMHEAFRGKITALVALGFTNNKNDTNGHGT
VAGSVLGNAGTGNAPQANLVQSIMDSGGGLGPNLQTLSPQAFSAGARHTNS
WGAAVNGATYDTSRVDVYRKNDITLFAAGNRPNGTISAPGTAKNALIVGATEN
LRPSFGSADNHNHVAQSFESRGTDKRIKPDWMAPTIILSARSSLPDSSFWANHG
SKYAYMGTSMATPIVAGNVAQRFKRNIGITPKPSLLKALIAAGADYGLGYPNG
NOGWGRVTDKSLNWAYNESALSTSKATYTFATAGKPLKISLVMSDAPASTAS
VTILNDLDIVITAPNGTRVVGNDFSAPEDNNWGRNVENVFINSQSGTGTIEVQAY
NVPVGPQNFSLAVN"
ORIGIN
Alignment Scores:
Pred. No.: 2,46e-117 Length: 1920
Score: 1998.50 Matches: 382
Percent Similarity: 94.01% Conservative: 26
Best Local Similarity: 88.02% Mismatches: 25
Query Match: 89.31% Indels: 1
DB: Gaps: 1
US-09-985-689A-3 (1-433) x AB046403 (1-1920)
Qy 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnAsnAsnTyrGlyLeuTyr 20
Db 616 AATGATGTGCCAGAGGATTTGTCAAAGCGGATGTGGCAGACGACTACGGTTTGAT 675
Qy 21 GlyGlnGlyGlnValValAlaAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 676 GGCAAGGCCAGATTTGCGAGTTGCGATACCTGATGATGGATACAGAAAGAACGACAGT 735
Qy 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 736 TCGATGATGAGACCTTCGCGGTGTAATAATACAGCACATATATGCACTGGGTGGAGCAT 795
Qy 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn--- 79
Db 796 AATCGGAATGATACGAACGGTTCATGTTACCATCTGCAACCTTATTCAGCCCAAGCATTCAGT 855
Qy 80 AlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 99
Db 856 GCAACGAATAAGGAATGGCACCTCAAGCGAATCTGTTTTCATTCATCATCATGATGAC 915
Qy 100 SerGlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaIleTrpAsn 119
Db 916 AGTGTGGCTTGGAGGCTTGCCTTCATCTGCAACCTTATTCAGCCCAAGCATTCAGT 975
Qy 120 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAlaGlnGlyAlaIleThrAla 139
Db 976 GCAGGTGCCAGAAATTCATACAACTCTGGGGGGCAGCGGTGAATGGGCTTACAGACA 1035
Qy 140 AsnSerArgGlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAla 159
Db 1036 GATTCAGAAATGTGGATGACTATGTAAGGAAAATGATATGACGATCTCTTTTCGGGCT 1095
Qy 160 GlyAsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 179
Db 1096 GGGGAATGAAAGCCGACGCGGTACCATCATGTCCTGGTACGCTGTAAGGAAAACGCGCAT 1155
Qy 180 ThrValGlyAlaThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsnProAsn 199
Db 1156 ACAGTGGCGGACCGAANAACCTGGTCCAGCTTCGGTTCCTATGAGATATATTAAC 1215
Qy 200 HisIleAlaGlnPheSerArgGlyAlaThrArgAspGlyArgIleLysProAspVal 219
Db 1216 CACGTTGCACAGTTCTCTTCCTCCGTCGCGCCGACAAAGATGGCGAAATCAAGCCTGATGC 1275
Qy 220 ThrAlaProGlyThrPheIleLeuSerAlaArgSerLeuAlaProAspSerSerPhe 239
```

```
Db 1276 ATGGCGCCAGGACATACATTTTATCAGCAAGATCTCTCTTGCAACCCGATTCCTCTTC 1335
Qy 240 TrpAlaAsnTyrAsnSerLysTyrAlaIleTyrMetGlyGlyThrSerMetAlaThrProIle 259
Db 1336 TGGCGCAATCATCAGCAAGAAATGCTATATGGGTGGAAACGCTCCATGGCAACACCGATT 1395
Qy 260 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro 279
Db 1396 GTTGGCGGGAATCTGCACAGCTCCGTGAGCATTTTGTGAAAAATAGAGGAATCACTCCT 1455
Qy 280 LysProSerLeuLysAlaAlaLeuIleAlaGlyAlaThrAspValGlyLeuGlyTyr 299
Db 1456 AAGCCTTCCTATGTAAAGCAGCTTTGATTCGAGGTGCTGCTGATGTTGGATTGGATTAT 1515
Qy 300 ProSerGlyAspGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaIleTyr 319
Db 1516 CCGAACGAAACCAAGGATGGGCGGAGTGACCTGGATAAATCGTTGAACGTTGCCTAT 1575
Qy 320 ValAsnGluAlaThrAlaLeuThrThrGlyGlnLysAlaThrTyrSerPheGlnThrGln 339
Db 1576 GTGAACGAATCCAGTGCCCTTATCAACTAGCCAAAAGCGACATATACCTTTACTGCAACG 1635
Qy 340 AlaGlyLysProLeuLysIleSerLeuValTyrThrAspAlaProGlySerThrAla 359
Db 1636 GCGGCGAAGCATTTGAAATCTCCCTGATGTCGCGATGCCCTGCAACGACTACTGCT 1695
Qy 360 SerTyrThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyGlnLysTyr 379
Db 1696 TCTGTAACCTCGTCAATGATTTGGATTTGGTCAATACAGCAACCAACGGAACAGATAT 1755
Qy 380 ValGlyAsnAspPheSerTyrProTyrAspAsnAsnTyrAspGlyArgAsnAsnValGlu 399
Db 1756 GTGCGGAATGACTCTCAGCACCATTTGACAAATACCTGGGATGGCGCAATAACGTAGAA 1815
Qy 400 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaIleTyrAsn 419
Db 1816 AATGTATTTATTAATTCGCCCAAGTGGAACATATATACCATTCAGGTGCAAGCATATAAT 1875
Qy 420 ValProSerGlyProGlnArgPheSerLeuAlaIleValHis 433
Db 1876 GTGCGGTGGACCAAAACTTCTCGTTGGCAATTTGTGAAC 1917
RESULT 10
LOCUS AR368116 1920 bp DNA linear PAT 12-SEP-2003
DEFINITION Sequence 3 from patent US 6376227.
ACCESSION AR368116
VERSION AR368116.1 GI:34601777
KEYWORDS SOURCE
ORGANISM Unknown.
REFERENCE Unclassified.
AUTHORS 1 (bases 1 to 1920)
TAKAIWA,M., OKUDA,M., SAEKI,K., KUBOTA,H., HITOMI,J., KAGEYAMA,Y.,
SHIKATA,S. and NOMURA,M.
TITLE Alkaline protease
JOURNAL Patent: US 6376227-A 3 23-APR-2002;
FEATURES Location/Qualifiers
1..1920
source /organism="unknown"
/mol_type="genomic DNA"
ORIGIN
Alignment Scores:
Pred. No.: 2,46e-117 Length: 1920
Score: 1998.50 Matches: 382
Percent Similarity: 94.01% Conservative: 26
Best Local Similarity: 88.02% Mismatches: 25
Query Match: 89.31% Indels: 1
DB: Gaps: 1
US-09-985-689A-3 (1-433) x AR368116 (1-1920)
```

QY 1 AsnAspValAlaArgGlyLeuValLysAlaAspValAlaGlnAsnAsnTyrGlyLeuTyr 20
Db 616 AATGATGTGGCCAGAGGTATTGTCAAAGCGGATGTGGCAGACAGCTACGGTTTGTAT 675
QY 21 GlyGlnGlyGlnValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 676 GGACAGGCCAGATTGTCCGAGTTGCCGATACTGGATTGGATACAGGAAGAACGACAGT 735
QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 736 TCGATGCATGAAGCCTTCGCGGTGAAAAATAACAGCAGCTATATGCACCTGGGTGCGACGAAT 795
QY 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn--- 79
Db 796 AATGCCAATGATACGAACGGTCATGGTACCATGTGGCAGGTTCGGTATTAGGAATGGC 855
QY 80 AlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 99
Db 856 GCAACGAATAAGGAATGGCACCTCAAGCGAATCTGGTTTTTCAATCCATCATGGATAGC 915
QY 100 SerGlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsn 119
Db 916 AGTGTGGGCTTGGAGGCTTGCCTTCCCAATCTGCAAACTTATTTCAGCCCAAGCAATTCAGT 975
QY 120 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyValaTyrThrAla 139
Db 976 GCAGGTGCCAGATTATACAACTCTTGGGGGCGAGCGGTGAATGGGGCTACACGACA 1035
QY 140 AsnSerArgGlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAla 159
Db 1036 GATTCAGAAATGTGGATGATATGTAAGGAAAAATGATATGACGATTCTTTTCGCGCT 1095
QY 160 GlyAsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 179
Db 1096 GGAATGAAGCCGACCGCGGTACCATCATGTGCCTGTGTACGGCTPAAAAAGCCATA 1155
QY 180 ThrValGlyAlaThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsnProAsn 199
Db 1156 ACAGTCGGCGCAACCGAAACCTGGGTCCAGCTTCGGTTCCGTATGCAGATAATATTAAAC 1215
QY 200 HisIleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspVal 219
Db 1216 CAGTTTGACAGTTCTTCTCCGTGGCGGACAAAGATGGCGGAATCAACGCTGATGTC 1275
QY 220 ThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerPhe 239
Db 1276 ATGGCGCCAGGACATACATTTTATCAGCAAGATCTTCTTTCACCCGATTCTCCCTC 1335
QY 240 TrpAlaAsnTyrAsnSerIlyAlaTyrMetGlyThrSerMetAlaThrProIle 259
Db 1336 TGGGCGAATCATGACAGCAAAATATGCTTATATGGGTGGAACGTCCTATGGCAACCCGATT 1395
QY 260 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro 279
Db 1396 GTTGGGGGAATGTTGCACAGCTCCGTGAGCATTITGTGMAAAATAGAGGAATCACTCT 1455
QY 280 LysProSerLeuIleLysAlaAlaLeuIleAlaGlyAlaThrAspValGlyLeuGlyTyr 299
Db 1456 AAGCCTTCCCTATTGAAGCAGCTTGTATGTCAGTGTCTGTGATGTTGGATTGGGTTAT 1515
QY 300 ProSerGlyAspGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 319
Db 1516 CCGAACGGAACCAAGGATGGGGCGAGTGACCTGGATTAATCTTGAAGTTGCCCTAT 1575
QY 320 ValAsnGluAlaThrAlaLeuThrThrGlyGlnLysAlaThrTyrSerPheGlnThrGln 339
Db 1576 GTGAACGAATCCAGTCCCTATCACTAGCCAAAAAGCGACATAACCTTACTGCAACG 1635
QY 340 AlaGlyLysProLeuLysIleSerLeuValTrpThrAspAlaProGlySerThrThrAla 359
Db 1636 GCGGGCAAGCCATTGAAATCTCCCTGGTATGGTGGATGGCCCTGCAGCACTACTGCT 1695

QY 360 SerTyrThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyGlnLysTyr 379
Db 1696 TCTGTAAACCTCGTCAATGATTGGATTGGTCATTACAGCACCACCAACGAAACAAGATAT 1755
QY 380 ValGlyAsnAspPheSerTyrProTyrAspAsnAsnTrpAspGlyArgAsnAsnValGlu 399
Db 1756 GTCCGGAATGACTTCTCAGCACCATTTGACATACTCGGATGGCGCAATAACGTAGAA 1815
QY 400 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 419
Db 1816 AATGTATTATTAAATTCGCCCAAAATGGAACATATACCATTSAGGTGCAAGCATATAT 1875
QY 420 ValProSerGlyProGlnArgPheSerLeuAlaIleValHis 433
Db 1876 GTGCCGTTGGACCACAAAACCTTCCTGTTGGCAATTGTGAAC 1917
RESULT 11
AR368118 1923 bp DNA linear PAT 12-SEP-2003
LOCUS Sequence 7 from patent US 6376227.
DEFINITION AR368118
ACCESSION AR368118
VERSION AR368118.1 GI:34601779
KEYWORDS Unknown.
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1923)
AUTHORS Takaiwa,M., Okuda,M., Saeki,K., Kubota,H., Hitomi,J., Kageyama,Y.,
Shikata,S. and Nomura,M.
TITLE Alkaline protease
JOURNAL Patent: US 6376227-A 7 23-APR-2002;
FEATURES Location/Qualifiers
source 1..1923
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN
Alignment Scores:
Pred. No.: 3,82e-117 Length: 1923
Score: 1995.50 Matches: 380
Percent Similarity: 94.01% Conservative: 28
Best Local Similarity: 87.56% Mismatches: 25
Query Match: 88.18% Indels: 1
Gaps: 6
US-09-985-689A-3 (1-433) x AR368118 (1-1923)
QY 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnAsnAsnTyrGlyLeuTyr 20
Db 619 AATGATGTTCGCGCTGGAAATGTCAAGCGGATGTGCTCAGAGCAGCTACGGTTGTAT 678
QY 21 GlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 679 GGACAAAGGACAGATCGTAGCGTTGCCGATACAGGCTTGATACAGTCGCAATGACGT 738
QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 739 TCGATGCATGAACCTTCCCGGGAATAATTACTGCATTATATGCAATGGACGCGACGAT 798
QY 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn--- 79
Db 799 AATGCCAATCATACGAATGTCTATGTACGATGTGGCTGGCTCCGTATTAGGAACCGC 858
QY 80 AlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 99
Db 859 TCCACTTAATAAGGAATGGCGCTCAGCGAATCTTAGTCTTCCATCTATCATGGATAGC 918
QY 100 SerGlyGlyLeuGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsn 119
Db 919 GGTGGGGGACTTGGAGGACTTACCTTCGAATCTGCAACCTTATTTCAGCCCAAGCATACGT 978
QY 120 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyValaTyrThrAla 139

Db 979 GCTGGTCCAGATTTCATACAACTCTCGGGAGCAGCAGTGAATGGGGCTTACACACA 1038
Qy 140 AnSerArgGlnValAspGluTyrValArgAsnAspMetThrValLeuPheAlaAla 159
Db 1039 GATTCAGAAATGTGGATGACTATGTGCGCAAAATGATATGACCATCTTTTCGCTGCC 1098
Qy 160 GlyAsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaIleAsnAlaIle 179
Db 1099 GGAATGAAGGACCGAACCGCGGACCATCATGTGACCGGACAGCTTAATAATGCAATA 1158
Qy 180 ThrValGlyAlaThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsnProAsn 199
Db 1159 ACATGCGGAGCTACGGAAACCTCGCCCAAGCTTTGGGTCTTATGCGGCAATATCAAC 1218
Qy 200 HisIleAlaGlnPheSerArgGlyAlaThrArgAspGlyArgIleIleProAspVal 219
Db 1219 CATGTGGCAGTTCCTTCACGTGGACCGCAAGATGGGACGATCAACCGGATGTC 1278
Qy 220 ThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerPhe 239
Db 1279 ATGGCACCGGAACTTCATCTATCATACAGATCTTCTCTTGCACCGATTCTCTCTTC 1338
Qy 240 TrpAlaAsnTyrAsnSerIleTyrAlaTyrMetGlyThrSerMetAlaThrProIle 259
Db 1339 TGGCGCAACCATGACATGTAATATGATCATGATGGGTGGAACTCCATGGCTACACCGATC 1398
Qy 260 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro 279
Db 1399 GTTGTCTGGAAACGTGGCAGCTTCGTGAGCATTTTGTGAAAAACAGAGGATCACACCA 1458
Qy 280 LysProSerLeuIleLysAlaAlaLeuIleAlaGlyAlaThrAspValGlyLeuGlyTyr 299
Db 1459 AAGCCTTCTCTATTAAAGCGGCACCTGATTGCGGTGACAGCTGACATGGCCCTTGGCTAC 1518
Qy 300 ProSerGlyAspGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 319
Db 1519 CCGACGGTAACCAAGATGGGACGATGACATTTGGATTAATCTCTGACGTTCCCTAT 1578
Qy 320 ValAsnGluAlaThrAlaLeuThrThrGlyGlnLysAlaThrTyrSerPheGlnThrGln 339
Db 1579 GTGACGAGTCCAGTCTCTATCCACCGACCAAAAGCGACGTACTCGTTACTGCTACT 1638
Qy 340 AlaGlySerProLeuLysIleSerLeuValTyrThrAspAlaProGlySerThrThrAla 359
Db 1639 GCGCGAAGCCTTTGAAATCTCCCTGGTATGCTGATGCTGATGCTGCTGCTGCTGCTGCT 1698
Qy 360 SerTyrThrLeuValAsnAspLeuValIleThrAlaProAsnGlyGlnLysTyr 379
Db 1699 TCGGTAACTGTCATATGATCTGGACCTTGTCATTTACCGCTCCAAATGGCACACAGTAT 1758
Qy 380 ValGlyAsnAspPheSerTyrProTyrAspAsnAsnTyrAspGlyArgAsnAsnValGlu 399
Db 1759 GTAGGAATGACTTTACTTCGCCATCAATGATGATGATGATGATGATGATGATGATGATGAT 1818
Qy 400 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 419
Db 1819 AATGTAATTAATGACCAACAAAGCGGACGTATACAAATGAAGTACAGGCTTATAAC 1878
Qy 420 ValProSerGlyProGlnArgPheSerLeuAlaIleValHis 433
Db 1879 GTACCGGTGGACACAGAACTTCTCGTTGGCAATTTGTAAT 1920
RESULT 12
AX839476 1305 bp DNA linear PAT 15-DEC-2003
LOCUS AX839476
DEFINITION Sequence 2 from Patent EP1347044.
ACCESSION AX839476
VERSION AX839476.1 GI:39922766
KEYWORDS
SOURCE Bacillus sp. KSM-KP43
ORGANISM Bacillus sp. KSM-KP43
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
REFERENCE 1

AUTHORS Okuda, M.K., Sato, T.K., Saito, K.K., Sumitomo, N.K., Izawa, Y.K.,
Saeki, K.K., Kobayashi, T.K. and Nomura, M.K.
TITLE Alkaline protease
JOURNAL Patent: EP 1347044-A 2 24-SEP-2003;
Kao Corporation (JP)
FEATURES
source Location/Qualifiers
1. .1305
/organism="Bacillus sp. KSM-KP43"
/mol_type="unassigned DNA"
/db_xref="taxon:109322"
CDS
1. .1305
/notes="unnamed protein product"
/codon_start=1
/transl_table=11
/protein_id="CAB55146.1"
/db_xref="GI:39322767"
/translation="NDVARGIVKADVAOSSYGLYGGQIVAVADTGLDTGRNDSNMHE
AFGKITALYALGRITNNNDTHGHGTHVAGSVLNGSTNKGMAPQANLVFQSIMDSGG
GUGLPSNIQTLPFSQAYSAGARIHTNSWGAANGAYTTDSRVNDDTIRKNDMLIFAA
DNVNGGTISAPQAKNAITVTGATENLSPFSYADININVAQSSPGPTKDGRIKP
GNGPGTFLSARSLAPDSFWANHDSKYAVGGTSMATPIVAGNVAQLRHFVKNR
GITPKSLKALIAAGADIGLYPNGNCGWRVTLDKSLNAVYNSSLSSTSQAT
YSTPAGKPLKISLVWSDAPASTTASVTLVNDLIVITAPNGTQYVGNDFTSYNDN
WDRNNVNFVNAFQSGTYTIEVQAYNVPGPTFSLAVN"

ORIGIN

Alignment Scores:
Pred. No.: 2,79e-117 Length: 1305
Score: 1994.50 Matches: 380
Percent Similarity: 94.01% Conservative: 28
Best Local Similarity: 87.56% Mismatches: 25
Query Match: 88.14% Indels: 1
DB: Gaps: 1

US-09-985-689A-3 (1-433) x AX839476 (1-1305)

Qy 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnAsnAsnTyrGlyLeuTyr 20
Db 1 AATGATTTTCGGCTGGATTTCAAGCGGATGTGGCTCAGACGAGCTACGGGTGTAT 60
Qy 21 GlyGlnGlnValValAlaAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 61 GCACAGACACAGATCGTAGCGTTGCCGATACAGGCGTTGATACAGCTCGCAATGACAGT 120
Qy 41 SerMetHisGluAlaPheArgGlyIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 121 TCGATGATGAGCTTCCCGGGGAAATTTACTGCAATATATGATGATGATGATGATGATGAT 180
Qy 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn--- 79
Db 181 AATGCCAATGATACGATGCTCATGTGACGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 240
Qy 80 AlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 99
Db 241 TCCACTAATAAGAAATGGCGCTCAGCGGAACTTAGCTCTTCCAACTATCATGATGATGATGAT 300
Qy 100 SerGlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTyrAsn 119
Db 301 GGTGGGGGACTTGAGGACTACCTTCGAATCTGCAACCTTATTTCAGCCCAAGCATACAGT 360
Qy 120 AlaGlyAlaArgIleHisThrAsnSerTyrGlyAlaProValAsnGlyAlaTyrThrAla 139
Db 361 GCTGTGCCAGAAATTCACAACTCTCGGGGACGACGATGATGATGATGATGATGATGATGATGAT 420
Qy 140 AsnSerArgGlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAla 159
Db 421 GATTCAGAAATGTGGATGACTATGTGGCAAAATGATGATGATGATGATGATGATGATGATGAT 480
Qy 160 GlyAsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 179
Db 481 GGGAAATGAAGACCGAAGCGGGAACCATCATGTCACCGACGACGATGATGATGATGATGATGAT 540
Qy 180 ThrValGlyAlaThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsnProAsn 199

```

Db 541 ACAGTCGGAGTCACGAAACCTCCGCCCAAGCTTTGGGTCTTATGCGGCAATATCAAC 600
Qy 200 HisleAlaGlaPheSerArgGlyAlaThrArgAspGlyArgIleIleYProAspVal 219
Db 601 CATGTGGCAGTTCCTTCAGTGGACCGCAAAAGGATGGACGATCAAAACCGGATGTC 660
Qy 220 ThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerPhe 239
Db 661 ATGGCACCGGAACTTCATCTATCAGCAAGATCTCTCTTGACCGGATTCCTCTTC 720
Qy 240 TrpAlaAsnTyAsnSerIleYThrAlaTyMetGlyGlyThrSerMetAlaThrProIle 259
Db 721 TGGCGCAACATGACAGTAAATATGATCATGATGGTGGAACTCATGCTACACCGATC 780
Qy 260 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLeuAsnArgGlyIleThrPro 279
Db 781 GTTGCTGGAAACGTGGCAGCTTCGTGAGCATTTGTGAAAAACAGAGGATCACACCA 840
Qy 280 LysProSerLeuIleLysAlaLeuIleAlaGlyAlaThrAspValGlyLeuGlyTy 299
Db 841 AAGCCTCTCTATTAAAGCGCAGCTGATTCGCGGTGCAGCTGACATCGGCTTGCTTAC 900
Qy 300 ProSerGlyAspGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTy 319
Db 901 CCGAACGGTAACCAAGATGGGAGGAGTGCATTCGATTAATCCTGACGTTGCCCTAT 960
Qy 320 ValAsnGluAlaThrAlaLeuThrThrGlyGlnLysAlaThrTySerPheGlnThrGln 339
Db 961 GTGAACGAGTCCAGTTCTCTATCCACAGCAAAAGCGAGCTACTCGTTTACTGCTACT 1020
Qy 340 AlaGlyLysProLeuLysIleSerLeuValTrpThrAspAlaProGlySerThrThrAla 359
Db 1021 GCGCGCAAGCCTTTGAAATCTCCCTGGTATGCTGATGCGCCCTGCGAGCAAACTGCT 1080
Qy 360 SerTyThrLeuValAsnAspLeuValIleThrAlaProAsnGlyGlnLysTy 379
Db 1081 TCGTAAACGCTTGCAATGATCTGACCTTGTCATTACCGCTCCAAATGGCACACAGTAT 1140
Qy 380 ValGlyAsnAspPheSerTyProTyAspAsnAsnTrpAspGlyArgAsnAsnValGlu 399
Db 1141 GTAGGAATAGCTTTACTTCGCGCATACATGATGATGGATGGCGGCGCAATACGTAGAA 1200
Qy 400 AsnValPheIleAsnAlaProGlnSerGlyThrTyThrIleGluValGlnAlaTyAsn 419
Db 1201 AATGTTATTTAATATCCACCAAGCGGAGCTATACAAATGAGTGACAGCTTATTAAC 1260
Qy 420 ValProSerGlyProGlnArgPheSerLeuAlaIleValHis 433
Db 1261 GTACCGGTTGGACACAGACCTTCTCGTTGGCAATTTGTAAT 1302

```

RESULT 13

AB051423

LOCUS

AB051423

DEFINITION

Bacillus sp. KP43

ACCESSION

AB051423.2

VERSION

GI:20521154

KEYWORDS

SOURCE

Bacillus sp.

ORGANISM

Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

REFERENCE

1

Itoh, S. and Saeki, K.

AUTHORS

TITLE

new protease

JOURNAL

Unpublished

REFERENCE

2 (bases 1 to 1923)

AUTHORS

Saeki, K.

TITLE

Direct Submission

JOURNAL

Submitted (21-NOV-2000)

Katsuhisa Saeki, KAO, CORPORATION;

2606, AKABANE, ICHIKAIWACHI, HAGA, TOCHIGI 321-3486, Japan

(E-mail: 38718@kasanet.kao.co.jp, Tel: 81285687471 (ex. 7471),

Fax: 81285687403)

On May 9, 2002 this sequence version replaced gi:14164344.

COMMENT

FEATURES

source

1. .1923

/organism="Bacillus sp. KSM-KP43"

/mol_type="genomic DNA"

/strain="KP43"

/db_xref="taxon:109322"

1. .1923

/genes="PROF"

1. .1923

/genes="PROF"

/codon_start=1

/transl_table=11

/product="protease"

/protein_id="BAB55674.2"

/db_xref="GI:20521155"

translation="MEKKKKVFLSVLSAAAILSTVALSNPSAGGARNFOLDKFGIQT
 TDAKFSKOGTGAAPLVESENVLKPKLOKLETVPANLKLHIIOFNGPILEETKO
 QLEKTKGKILDIYIPDYAVIYEVDKVSATSTIEHVESVEPLYPIRIDPOLFTKAS
 ELKVALDTPKKNKEVQIRGEIQAFALSNLDLYITAKPEYKWNDAVARGIVADV
 AQSSYGLGGQIVAVADTGLDTRGNSDHEAFRQKITALTALGRTNANDINGHGT
 HVAGSVLNGSTNKGMAPQANLVFQISMDGGGLGSPNLQTLFQTSAGARAHITN
 SWGAAYNGAYTDSRVDDYVRKNDITLFAAGNEGPGGTISAPGTAKNAITVGTAE
 NLKPSVGSADININHVAFQSSRPTKDGRIKPDVMAFGFTILSARSLAPDSSFWANH
 DSKVAYMGGTSMATPIVAGNVAOLREHFVNKRGITPKSLKALAGAADIGLGPVN
 GNGWGRVTLDKSLINAVYNESLSLSOKATYSETATAGKPKIKISLVKSDAPASITA
 SVTLVNDLLDVIITAPNGTQYVGNDFTPSPNDNWDGRNVNFINAPQSGTYTIEVQA
 INVPVGPQTFLSLAIVN"

ORIGIN

Alignment Scores:

Pred. No.: 4,42e-117 Length: 1923
 Score: 1994.50 Matches: 380
 Percent Similarity: 94.01% Conservative: 28
 Best Local Similarity: 87.56% Mismatches: 25
 Query Match: 88.14% Indels: 1
 DB: 1 Gaps: 1

US-09-985-689A-3 (1-433) x AB051423 (1-1923)

Qy 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnAsnAsnTyGlyLeuTy 20

Db 619 AATGATGTTCCGCGTGAATTCGAAGCGGATGTGGCTCAGACGACGATCGGTTGTAT 678

Qy 21 GlyGlnGlyValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40

Db 679 GGACAGGACAGATCGTAGCGTTGCCGATACAGGCTTGATACAGTCGCAATGACAGT 738

Qy 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyThrAlaLeuGlyArgThrAsn 60

Db 739 TCGATGTCATGAACCTTCGCGGAAATTTACTGCAATATATGTCATTTGGACGACGAAAT 798

Qy 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn-- 79

Db 799 AATGCCAATGATACGAATGTCATGTCAGCATGTGGCTGGCTCCCTATTAGGAACGGC 858

Qy 80 AlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 99

Db 859 TCCACTAATAAGGAATGGCGCTCAGGCGAATCTAGTCTTCCAAATCTATCATGGATAGC 918

Qy 100 SerGlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTyAsn 119

Db 919 GGTGGGGGACTTGGAGGACTACCTTCGATCTGCAACCTTATTCAGCCAGCATACAGT 978

Qy 120 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyThrAla 139

Db 979 GCTGTGCCAGATTCATACAAACTCCTGGGAGCAGCAGTGAATGGGCTTACACAACA 1038

Qy 140 AsnSerArgGlnValAspGluTyValArgAsnAsnAspMetThrValLeuPheAlaAla 159

Db 1039 GATTCCAGAAATGTGATGACTATGCGCAAAATATGATATGACGACTCTTTTCGTGCC 1098

Qy 160 GlyAsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 179

1099 GGGAAATGAAGGACCGAAGCGGGAACCATCAGTCGACAGGACAGCTAAAAATGCAATA 1158
 180 ThrValGlyAlaThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsnProLeu 199
 1159 ACAGTCGGAGCTACGGAAACCTCCGCCCAAGCTTTGGGTCTTATGCGGACAATATCAAC 1218
 200 HisIleAlaGlnPheSerArgGlyAlaThrArgAspGlyValGlyLeuValAspVal 219
 1219 CATGTGGCAGATCTCTTCAGTCGACCGACAAAGATGGAGGATCAACCGGATGTC 1278
 220 ThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 239
 1279 ATGGCACCAGGAACTTATCTATCATCAGCAAGATCTCTTCGACCGGATTCCTCCCTTC 1338
 240 TrpAlaAsnTyrAsnSerIleValAlaTyrMetGlyGlyThrSerMetAlaThrProIle 259
 1339 TGGCGAACCATACAGATTAATATGATATCATGCTGGTGGAGCTCCATGGCTACACGATC 1398
 260 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLeuAsnArgGlyIleThrPro 279
 1399 GTTGTCTGGAACCTGGCAGACCTTCGTGAGCATTTTGTGAAAAACAGAGGATCAACCA 1458
 280 LysProSerLeuIleValAlaLeuIleAlaGlyAlaThrAspValGlyLeuGlyTyr 299
 1459 AAGCCTTCTTATTAAGCGGACCTGATTCGCGTGCAGCTGACATCGGCTTGGCTAC 1518
 300 ProSerGlyAspGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 319
 1519 CCGAACGGTAACCAAGGATGGGACGAGTGACATTCGATAAATCCCTGAACGTGGCTAT 1578
 320 ValAsnGluAlaThrAlaLeuThrThrGlyGlnLysAlaThrThrSerPheGlnThrGln 339
 1579 GTGAACGAGTCCAGTCTCTATCCACGACGCAAAAGCGACGACTCTGTTTACTGCTACT 1638
 340 AlaGlyLysProLeuLysIleSerLeuValTyrThrAspAlaProGlySerThrThrAla 359
 1639 GCGGCAAGCTTTGAAATCTCCCTGGTATGTCGTGATGTCGCTCGGACCAACTGCT 1698
 360 SerTyrThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyGlnLysTyr 379
 1699 TCCGTAAACGTTGCAATGATCTGGACCTTGTATTACCGCTCCAAATGGCACACAGTAT 1758
 380 ValGlyAsnAspPheSerTyrProTyrAspAsnAsnTyrAspGlyArgAsnAsnValGlu 399
 1759 GTAGGAATGACTTACTTCGCCATACATGATACTGGATGGCGGCAATACAGTAGAA 1818
 400 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 419
 1819 AATGTATTTATTAATGCACCAAGCGGACCTATATCAATGAGGTACAGGCTTATAAC 1878
 420 ValProSerGlyProGlnArgPheSerLeuAlaIleValHis 433
 1879 GTACCGGTGGACCAAGACCTTCTCTTGGCAATTTGTGAAT 1920

RESULT 14
 AR368117
 LOCUS 1923 bp DNA linear PAT 12-SEP-2003
 DEFINITION Sequence 5 from patent US 6376227.
 ACCESSION AR368117
 VERSION AR368117.1 GI:34601778

KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE 1 (bases 1 to 1923)
 AUTHORS Takaiwa, M., Okuda, M., Saeki, K., Kubota, H., Hitomi, J., Kageyama, Y.,
 Shikata, S., and Nomura, M.
 TITLE Alkaline protease
 JOURNAL Patent: US 6376227-A 5 23-APR-2002;
 FEATURES Location/Qualifiers
 1..1923
 /organism="unknown"
 /mol_type="genomic DNA"

ORIGIN

Alignment Scores:
 Pred. No.: 4.42e-117 Length: 1923
 Score: 1994.50 Matches: 380
 Percent Similarity: 94.01% Conservative: 28
 Best Local Similarity: 87.56% Mismatches: 25
 Query Match: 88.14% Indels: 1
 Gaps: 6

US-09-985-689A-3 (1-433) x AR368117 (1-1923)

QY 1 AsnAspValAlaAlaArgGlyIleValLysAlaAspValAlaGlnAsnAsnTyrGlyLeuTyr 20
 Db 619 AATGATGCTGCGGTGAAATTTGCAAGCGGATGTGCTCAGACGACGCTACGGGTTGTAT 678
 QY 21 GlyGlnGlyGlnValValAlaAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
 Db 679 GGACAAAGACAGATCTGAGGTGGTTCGCAATACAGGCTTGATACAGGTGCGATGACAGT 738
 QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
 Db 739 TCGATGATGAAGCTTCCCGGGAAATTTACTGCATTATATGATTATGCGGACGACGAT 798
 QY 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn--- 79
 Db 799 AATGCCAATGATACGAATGCTGATGGTACGATGTGCTCGCTCCGTTATTAGGAACGGC 858
 QY 80 AlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 99
 Db 859 TCCACTAATAAGGAATGGCGCTCAGCGCAATCTAGCTTCCCAATCTATCATGATGATG 918
 QY 100 SerGlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTyrAsn 119
 Db 919 GGTGGGACCTGTGAGGACCTACCTTCGAATCTGCAACCTTATTTCAGCAAGCATACAGT 978
 QY 120 AlaGlyAlaArgIleHisThrAsnSerTyrGlyAlaProValAsnGlyAlaTyrThrAla 139
 Db 979 GCTGGTGCAGAAATTCATACAAACTCTCTGGGAGCAGCAGTGAATGGGCTTACACAACA 1038
 QY 140 AsnSerArgGlnValAlaAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAla 159
 Db 1039 GATTCCGAAATGTGGATGATCTATGTGCGCAAAATGATATGACGATCCTTTTCGCTGCC 1098
 QY 160 GlyAsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 179
 Db 1099 GGGAAATGAAGACCGAAGCGGGAACCATCATGTCACACGACAGCTAAAAATGCAATA 1158
 QY 180 ThrValGlyAlaThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsnProAsn 199
 Db 1159 ACAGTCGGAGCTACGGAAACCTCCGCCCAAGCTTTGGGTCTTATGCGGACAAATATCAAC 1218
 QY 200 HisIleAlaGlnPheSerArgGlyAlaThrArgAspGlyArgIleLysProAspVal 219
 Db 1219 CATGTGGCAGATCTCTTCAGTCGACCGCAAGGATGGACGATCAACCGGATGTC 1278
 QY 220 ThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 239
 Db 1279 ATGGCACCAGGAAACCTTCTATCATCAGCAAGATCTCTTCGACCGGATTCCTCCCTTC 1338
 QY 240 TrpAlaAsnTyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 259
 Db 1339 TGGCGAACCATGACAGATTAATATGATATCATGCTGGTGGAGCTCCATGGCTACACGATC 1398
 QY 260 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro 279
 Db 1399 GTTGTCTGGAACCTGGCAGACCTTCGTGAGCATTTTGTGAAAAACAGAGGATCAACCA 1458
 QY 280 LysProSerLeuIleLysAlaLeuIleAlaGlyAlaThrAspValGlyLeuGlyTyr 299
 Db 1459 AAGCCTTCTTATTAAGCGGACCTGATTCGCGTGCAGCTGACATCGGCTTGGCTAC 1518
 QY 300 ProSerGlyAspGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 319

Search completed: April 4, 2004, 08:06:52
Job time : 3595.27 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus p2n model

Run on: April 3, 2004, 16:37:54 ; Search time 335.908 Seconds
(without alignments)

5476.111 Million cell updates/sec

Title: US-09-985-689A-3

Perfect score: 2263

Sequence: 1 NDVARGIVKADVAQNNYGLY.....EVQAYNVPSGQRFSLAIVH 433

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame_plus_p2n.model -DEV=xlh
-Q/cgn2_1/USPTO_spool/US0985689/runat_31032004_161806_4145/app_query.fasta_1_3498
-DS=N Geneseq 29Jan04 -QFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MTN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pct -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US0985689 @CGN1_1_1238/runat_31032004_161806_4145 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N Geneseq 29Jan04: *
1: geneseqn1980s: *
2: geneseqn1990s: *
3: geneseqn2000s: *
4: geneseqn2001as: *
5: geneseqn2001bs: *
6: geneseqn2002s: *
7: geneseqn2003as: *
8: geneseqn2003bs: *
9: geneseqn2003cs: *
10: geneseqn2004s: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|----|--------------------|
| 1 | 2239 | 98.9 | 1299 | 2 | AAQ27516 Alkali-pr |
| 2 | 2040 | 90.1 | 3003 | 2 | AAV82382 Bacillus |
| 3 | 1988.5 | 88.3 | 1920 | 2 | AAx37277 Bacillus |
| 4 | 1995.5 | 88.2 | 1923 | 2 | AAx37279 Bacillus |
| 5 | 1994.5 | 88.1 | 1923 | 2 | AAx37278 Bacillus |
| 6 | 1994.5 | 88.1 | 1923 | 2 | AAx37278 Bacillus |
| 7 | 443 | 19.6 | 1977 | 2 | AAT85667 Thermococ |
| 8 | 422.5 | 18.7 | 1236 | 2 | AAx05926 WO9856926 |
| | | | | | AAx05920 Hyperther |

| | | | | | |
|----|-------|------|--------|---|--------------------|
| 9 | 422.5 | 18.7 | 1566 | 2 | AAT85668 |
| 10 | 422.5 | 18.7 | 1962 | 2 | AAT85695 |
| 11 | 422.5 | 18.7 | 1962 | 2 | AAX05929 Hyperther |
| 12 | 401.5 | 17.7 | 1977 | 2 | AAT85669 |
| 13 | 370.5 | 16.4 | 2121 | 4 | ABL54900 T. yonsei |
| 14 | 346.5 | 15.3 | 135838 | 7 | ABX34289 S. atrool |
| 15 | 343.5 | 15.2 | 2539 | 2 | AAT61454 Streptomy |
| 16 | 343.5 | 15.2 | 2809 | 2 | AAT61455 DhpA-mel |
| 17 | 336 | 14.8 | 1329 | 6 | ABK74643 Bacillus |
| 18 | 312 | 13.8 | 2532 | 2 | AAQ29134 Encodes R |
| 19 | 306.5 | 13.5 | 1859 | 2 | AAT85677 Thermococ |
| 20 | 305 | 13.5 | 2835 | 2 | AAT08141 Hyperther |
| 21 | 305 | 13.5 | 4765 | 2 | AAT08132 Protease |
| 22 | 305 | 13.5 | 4765 | 2 | AAT85670 Pyrococcu |
| 23 | 305 | 13.5 | 4765 | 2 | AAX05921 WO9856926 |
| 24 | 296.5 | 13.1 | 3743 | 9 | ADD24901 DNA encod |
| 25 | 295 | 13.0 | 1560 | 7 | AB237569 Streptomy |
| 26 | 295 | 13.0 | 59816 | 7 | AB237516 Streptomy |
| 27 | 295 | 13.0 | 59816 | 7 | AB237515 Streptomy |
| 28 | 282 | 12.5 | 564 | 2 | AAT08134 DNA seque |
| 29 | 282 | 12.5 | 564 | 2 | AAT85676 Thermococ |
| 30 | 281.5 | 12.4 | 898 | 2 | AAT08131 Hyperther |
| 31 | 281.5 | 12.4 | 2166 | 6 | AAD34433 C-termina |
| 32 | 281.5 | 12.4 | 2166 | 6 | ABL40498 DNA const |
| 33 | 279.5 | 12.4 | 546 | 4 | ABL53453 T. yonsei |
| 34 | 272.5 | 12.0 | 4716 | 9 | ABQ80437 CapA codi |
| 35 | 272.5 | 12.0 | 4740 | 7 | AB258957 Group B S |
| 36 | 272 | 12.0 | 3413 | 2 | AAV72330 F. balust |
| 37 | 270 | 11.9 | 807 | 2 | AAQ11835 Carbonyl |
| 38 | 269 | 11.9 | 810 | 7 | AB221470 Mature Ba |
| 39 | 269 | 11.9 | 866 | 2 | AAQ30104 Synthetic |
| 40 | 269 | 11.9 | 866 | 2 | AAQ30146 Synthetic |
| 41 | 269 | 11.9 | 866 | 2 | AAQ31697 Subtilisi |
| 42 | 269 | 11.9 | 866 | 2 | AAQ26410 Subtilisi |
| 43 | 269 | 11.9 | 1140 | 2 | AAT37502 Bacillus |
| 44 | 269 | 11.9 | 1140 | 2 | AAT37502 Bacillus |
| 45 | 269 | 11.9 | 1140 | 2 | AAV69598 B. lentus |

ALIGNMENTS

RESULT 1

AAQ27516
ID AAQ27516 standard; DNA; 1299 BP.

XX AC AAQ27516;
XX DT 05-FEB-1993 (first entry)
XX DE Alkali-protease Ya enzyme gene.

XX KW Alkali resistance; surface active agent resistance; detergent improver;
XX OS ss.
XX OS Bacillus sp. Y.

XX FH Key Location/Qualifiers
XX FT CDS 1..1299
XX FT /*tag= a

XX PN JP04197182-A.

XX PD 16-JUL-1992.

XX PF 28-NOV-1990; 90JP-00327110.

XX PR 28-NOV-1990; 90JP-00327110.

XX PA (LION) LION CORP.

XX DR WPI; 1992-288440/35.

XX DR P-ESDB; AAR26274.

PT DNA coding alkali-protease Ya enzyme - has good alkali and surfactant
 PT resistance and improves detergency.

XX Claim 3; Page 2; 17pp; Japanese.

XX The sequence is that of the alkali-protease Ya enzyme gene which can be
 CC used in the recombinant production of Ya enzyme. Ya enzyme is excellent
 CC in alkali resistance and surface active agent resistance and improves
 CC detergency

XX Sequence 1299 BP; 401 A; 246 C; 297 G; 355 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 2,27e-168 Length: 1299
 Score: 2339.00 Matches: 428
 Percent Similarity: 99.31% Conservative: 2
 Best Local Similarity: 98.85% Mismatches: 3
 Query Match: 98.94% Indels: 0
 DB: 2 Gaps: 0

US-09-985-689A-3 (1-433) x AAQ27516 (1-1299)

| | | | |
|----|-----|---|-----|
| Qy | 1 | AsnAspValAlaArgGlyLeuValAlaAspValAlaGlnAsnAsnTyrGlyLeuTyr | 20 |
| Db | 1 | ATGAGTGTAGCAAGAGGATAGTAAAGCTGATGTTGCAACAAACATACGGATTATAT | 60 |
| Qy | 21 | GlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer | 40 |
| Db | 61 | GGCAAGGTCACACTAGTTGCGATAGCGGACACAGGCTTAGATACAGGTCGTAAACATAGT | 120 |
| Qy | 41 | SerMetHisGluAlaPheArgGlyLeuValAlaLeuTyrAlaLeuGlyValArgThrAsn | 60 |
| Db | 121 | TCATGATGAAGCAATTCGCGGGAATAACACAGCTCTTTACGGTTAGAGAACTAAT | 180 |
| Qy | 61 | AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnAla | 80 |
| Db | 181 | AATGCGAGTGATCCGAATGGCATGGGCACACATGTAGCAGGTTCTGTACTTGGTAATGT | 240 |
| Qy | 81 | LeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSerSer | 100 |
| Db | 241 | TTAAATAAAGGAATGGCTCCGCAAGCTTAACCTTATGTTTCCAACTTATATGATAGCAGC | 300 |
| Qy | 101 | GlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsnAla | 120 |
| Db | 301 | GGAGGATTAGTGGCTTACCATCGAATTAATACGTTATTTAGTCAAGCTTGGAAATGCT | 360 |
| Qy | 121 | GlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaThrAlaAsn | 140 |
| Db | 361 | GGAGCAAGAATTATACATACTACTCTTGGGAGCCCCAGTAAATGGAGCGTACACTGCTAAC | 420 |
| Qy | 141 | SerArgGlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAlaGly | 160 |
| Db | 421 | TCGAGACAGTGGATGATGATGTTCCGAATAATGATATGACGGTACTTTTTCAGCTGT | 480 |
| Qy | 161 | AsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAlaIleThr | 180 |
| Db | 481 | AATGAAGGTCCTAATTCAGGAACAATTAGTGTCCAGGTACAGCGAATAATGCTATTACG | 540 |
| Qy | 181 | ValGlyValThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsnProAsnHis | 200 |
| Db | 541 | GTCGGCGACAGGAAACTATCGCCCAAGCTTCGGTTCGATAGCAGATAACCAATCAT | 600 |
| Qy | 201 | IleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspValThr | 220 |
| Db | 601 | ATTGCACAAATTTTCATCGAGAGGAGCTACGAGGATGCGAATTAAGCCCTGACGTAACA | 660 |
| Qy | 221 | AlaProGlyThrPheIleLeuSerAlaArgSerLeuAlaProAspSerSerPheTrp | 240 |
| Db | 661 | GCTCCTGGAACATTTATTTATCAGCAGCTTCTTCTTAGCTCCAGCTCTTCTGTTTGG | 720 |
| Qy | 241 | AlaAsnTyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleVal | 260 |
| Db | 721 | CGCAATTATACAGTAATAACGCGTATATGGCGGCTACCTCCATCGGCGACACCTATTGTT | 780 |

| | | | |
|----------|---|--|------|
| Qy | 261 | AlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrProLys | 280 |
| Db | 781 | GCAGGGAAATGTCGCGCAATACGTCAGCATTTTATAAAAAATAGAGGTATTACTCCTAAG | 840 |
| Qy | 281 | ProSerLeuIleLysAlaAlaLeuIleAlaGlyAlaThrAspValGlyLeuGlyTyrPro | 300 |
| Db | 841 | CCTTCTTTAATAAAGCTGACATTATCGCTGGTGCTACTGATGTTGGTTAGGATATCCT | 900 |
| Qy | 301 | SerGlyAspGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyrVal | 320 |
| Db | 901 | AGTGGTGACCAAGCTGGGGGCGTCTCTCTAGATAAATCGTTAAATGTCAGGTATGTC | 960 |
| Qy | 321 | AsnGluAlaThrAlaLeuThrThrGlyGlnLysAlaThrTyrSerPheGlnThrGlnAla | 340 |
| Db | 961 | AATGAAGCAACTGCATTAGCCACAGACAAAAAGCAACCTATTCGTTCCAAAGCACAAGCG | 1020 |
| Qy | 341 | GlyLysProLeuLysIleSerLeuValTyrThrAspAlaProGlySerThrThrAlaSer | 360 |
| Db | 1021 | GGTAACCTTTAAANAATCTCGTTAGTATGGACAGATGCTCTTGGAAAGTACACTGCATCT | 1080 |
| Qy | 361 | TyrThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyGlnLysTyrVal | 380 |
| Db | 1081 | TATACACTAGTTAATGATTTAGATCTAGTTATTACTCTCCGAATGACACAAAAATATGTA | 1140 |
| Qy | 381 | GlyAsnAspPheSerTyrProTyrAspAsnAsnTrpAspGlyArgAsnAsnValGluAsn | 400 |
| Db | 1141 | GGAAATGATTTTATGTTATCTTATGATATATTAATCTGGATGGTGGCAACAATGTTGAGAAC | 1200 |
| Qy | 401 | ValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsnVal | 420 |
| Db | 1201 | GTAITTTATAACGCTCCGCAATCTGGAACTATATAATTAATGAGGTTCAAGCGTATAATGA | 1260 |
| Qy | 421 | ProSerGlyProGlnArgPheSerLeuAlaIleValHis | 433 |
| Db | 1261 | CCATCTGGCCCAAGCGTTTCTCCTAGCTATGCTACAT | 1299 |
| RESULT 2 | | | |
| ID | AAV82382 standard; DNA; 3003 BP. | | |
| XX | AAV82382; | | |
| DT | 12-APR-1999 (first entry) | | |
| XX | Bacillus Jp170 protease gene. | | |
| XX | Protease; detergent; surfactant; leather processing; debittering; | | |
| KW | flavour; ss. | | |
| OS | Bacillus sp. | | |
| PH | Location/Qualifiers | | |
| FT | 846..2771 | | |
| FT | /*tag= a | | |
| FT | sig_peptide 846..944 | | |
| FT | /*tag= b | | |
| FT | mat_peptide 1470..2768 | | |
| FT | /*tag= c | | |
| XX | W09856927-A2. | | |
| XX | 17-DEC-1998. | | |
| XX | 09-JUN-1998; 98WO-US012005. | | |
| XX | 12-JUN-1997; 97US-00873479. | | |
| XX | (NOVO) NOVO NORDISK BIOTECH INC. | | |
| XX | Sloma A, Christianson L; | | |
| XX | WPI; 1999-080908/07. | | |
| DR | | | |


```

07-OCT-1998;   98NO-JP004528.
XX
XX
PR PR          97JP-00274570.
XX
XX PA          (KAOS ) KAO CORP.
XX
XX PI Takaiwa M, Okuda M, Saeki K, Kubota H, Hitomi J, Kageyama Y;
PI Shikata S, Nomura M;
DR WPI; 1999-287736/27.
XX P-PSDB; AAV17087, AAY17089.
XX Alkali protease from Bacillus used in washing powders.
XX
XX PS Disclosure; Page 53-58; 7lpp; Japanese.
XX
XX The invention relates to alkaline proteases produced by strains of
CC Bacillus. The proteases ability to digest casein is not inhibited by
CC oleic acid and they have a high stability to oxidising agents. The
CC alkaline protease of the invention has the following properties: (a) it
CC is active over the pH range 4-13 and has at least 80% of its optimum
CC activity over the range pH 6-12; (b) after 30 minutes at 40 deg. C it is
CC stable over the pH range 6-11; c) its isoelectric point is 8.9-9.1; (d)
CC its ability to digest casein is not inhibited by oleic acid; (e) it has
CC molecular weight about 43,000 by SDS-PAGE. The alkaline proteases can
CC used as enzymes in washing compositions for use in automatic dishwashers
CC and for washing clothes. The stability to oxidising agents allows the
CC enzyme to be an effective component of washing compositions including
CC bleaches. The present sequence represents an alkaline protease encoding
CC DNA. (Updated on 20-MAR-2003 to correct DR field.)
XX
XX SQ Sequence 1920 BP; 576 A; 411 C; 473 G; 460 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:      4,35e-149      Length:      1920
Score:          1998.50        Matches:     382
Percent Similarity: 94.01%    Conservative: 26
Best Local Similarity: 88.02% Mismatches:     25
Query Match:      88.31%     Indels:       1
DB:               2         Gaps:         1

US-09-985-689A-3 (1-433) x AAX37277 (1-1920)

Qy 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnAsnAsenTyrGlyLeuTyr 20
Db 616 AATCATGTGGCAGAGGTATTGTCAAAGCGGATGTGCACAGACGACGTACGGTTTGAT 675
Qy 21 GlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 676 GGACAAGCCAGATTTGCCAGTTGCCGACTCTGGATTCGATACAGAGAAGAACGACAGT 735
Qy 41 SerMetHisGlualaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 736 TCGATGCATGAAGCCCTTCGGCGGTAAATAACACACACTATATGCACTGGGTTCGACCAAT 795
Qy 61 AsnAlaAsnAppProAsnGlyHisGlyThrisValalaglySerValLeuGlyAsn--- 79
Db 796 AATGCGAATGATACCAACGGTCATGGTACCCTATGGCAGGTTCCGTTATTAGGAATGGC 855
Qy 80 AlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 99
Db 856 GCAACGAATAAGGAATGGCACCTCAAGCGAATCTGTTTTTCAATCCATCATGGATAGC 915
Qy 100 SerGlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsn 119
Db 916 AGTGGTGGGCTTGGAGGCTTCCCTTCCAATTCGAAACCTTATTTCAGCCAAGCAATTCAGT 975
Qy 120 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyValaTrThrAla 139
Db 976 CGAGGTGCCAGAAITCATACAAACTCTTGGGGGGCAGCGGTGAATGGGCCCTACACGACA 1035
Qy 140 AsnSerArgGlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAla 159

```

XX WO9918218-A1.
XX 15-APR-1999.
XX 07-OCT-1998; 98WO-JP004528.
XX 07-OCT-1997; 97JP-00274570.
XX (KAOS) KAO CORP.
XX Takaiwa M, Okuda M, Saeki K, Kubota H, Hitomi J, Kageyama Y;
XX Shikata S, Nomura M;
XX WPI; 1999-287736/27.
XX P-PSDB; AA17091.
XX Alkali protease from Bacillus used in washing powders.
XX Disclosure; Page 63-68; 71pp; Japanese.
XX The invention relates to alkaline proteases produced by strains of
XX Bacillus. The proteases ability to digest casein is not inhibited by
XX oleic acid and they have a high stability to oxidising agents. The
XX alkaline protease of the invention has the following properties: (a) it
XX is active over the pH range 4-13 and has at least 80% of its optimum
XX activity over the pH range 6-12; (b) after 30 minutes at 40 deg. C it is
XX stable over the pH range 6-11; c) its isoelectric point is 8.9-9.1; (d)
XX its ability to digest casein is not inhibited by oleic acid; (e) it has
XX molecular weight about 43,000 by SDS-PAGE. The alkaline proteases can be
XX used as enzymes in washing compositions for use in automatic dishwashers
XX and for washing clothes. The stability to oxidising agents allows the
XX enzyme to be an effective component of washing compositions including
XX bleaches. The present sequence represents an alkaline protease encoding
XX DNA. (Updated on 20-MAR-2003 to correct DR field.)
XX SQ Sequence 1923 BP; 578 A; 417 C; 474 G; 454 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 7.55e-149 Length: 1923
Score: 1995.50 Matches: 380
Percent Similarity: 94.01% Conservative: 28
Best Local Similarity: 87.56% Mismatches: 25
Query Match: 88.18% Indels: 1
DB: Gaps: 1

US-09-985-689A-3 (1-433) x AAX37279 (1-1923)

QY 1 AsnAspValAlaArgGlyIleVallyAlaAspValAlaGlnAsnAsnTyrGlyLeuTyr 20
Db 619 AATGATGTTGCGGTGGAATTTGCAAGCGGATGGCTCAGACAGCTACGGTTGTAT 678

QY 21 GlyGlnGlnValValAlaAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 679 GGACAAAGGACAGATCGTACGGTTGCCGATCAGGCGCTTGATACAGGTGCAATGACAGT 738

QY 41 SerMetHisGluAlaPheArgGlyIleThrAlaLeuTyrAlaLeuGlyValGthrAsn 60
Db 739 TCGATGCATGAGCCTTCGCGGGAAATTAATCTGCATTATATGCTTGGACGCGAAT 798

QY 61 AsnAlaAspAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn--- 79
Db 799 AATGCCAATGATACGAATGGTTCATGTGTACGCATGTGGCTGGCTCCGCTATTAGGAACCGC 858

QY 80 AlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 99
Db 859 TCCACTAATAAAGGAATGGCGCTCAGGGGAATCTAGTCTTCAATCTATCATGTATAGC 918

QY 100 SerGlyGlyLeuGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsn 119
Db 919 GGTGGGGGACTTGGAGGACTACCTTCGAATCTGCAAACTTATTGAGCCAGCATACAGT 978

QY 120 AlaGlyAlaArgIleHisThrAsnSerTyrGlyAlaProValAsnGlyAlaTyrThrAla 139

Db 979 GCTGGTGGCAGATTCATCAAACTCTCTGGGAGCAGCAGTGAATGGGCTTACACAACA 1038
QY 140 AsnSerArgGlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAla 159
Db 1039 GATTTCAGAAATGTGATGATGATGCGCAAAATGATGATGATGATGATGATGATGATG 1098
QY 160 GlyAsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 179
Db 1099 GGGATGAGGACCGAAGCGGGAACCATCAGTGCACAGGACAGCAGTAAATGCAATA 1158
QY 180 ThrValGlyAlaThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsnProAsn 199
Db 1159 ACAGTGGAGCTACGGAAACCTCCGCCCAAGCTTTGGTCTTATCGGACCAATATCAAC 1218
QY 200 HisIleAlaGlnPheSerArgGlyAlaThrArgAspGlyArgIleLysProAspVal 219
Db 1219 CATGTGCACAGTCTCTTCACGTGCACCGCAAGCATGGACGATCAACCGGATGTC 1278
QY 220 ThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 239
Db 1279 ATGGCACCGGAAACGTTTCTATCATCAGCAAGATCTTCTCTGACCGGATTCCTCCTTC 1338
QY 240 TrpAlaAsnTyrAsnSerIleValAlaTyrMetGlyGlyThrSerMetAlaThrProIle 259
Db 1339 TGGCGAAGCATCAGATTAATTAATGATCATCATGGTGGAACTGCTCATGGCTACACGATC 1398
QY 260 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro 279
Db 1399 GTTGTGMAAACGTGGCACAGCTTCGTGAGCATTTGTGMAAAACAGAGGCATCACACCA 1458
QY 280 LysProSerIleLysAlaAlaLeuIleAlaGlyAlaThrAspValGlyLeuGlyTyr 299
Db 1459 AAGCTTCTCTATTAAAGCGGCACATGATTCGCGGTGACGTACATCGGCTTGCTAC 1518
QY 300 ProSerGlyAspGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 319
Db 1519 CCGACGGAACCAAGGATGGGACGAGTGACATTCGATAAATCCTGACGTTGCTAT 1578
QY 320 ValAsnGluAlaThrAlaLeuThrThrGlyGlnLysAlaThrTyrSerPheGlnThrGln 339
Db 1579 GTGAACGAGTCCAGTTCTCTATCCACGCAAAAGCGACGCTACTCGTTTACTGCTACT 1638
QY 340 AlaGlyLysProLeuLysIleSerLeuValTyrThrAspAlaProGlySerThrThrAla 359
Db 1639 GCGGCGAGCGCTTGAATAATCTCCCTGGTATGCTGATGCCCTCGGAGCACACTGCT 1698
QY 360 SerTyrThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyGlnLysTyr 379
Db 1699 TCCGTAAACGCTTGTCAATGATCTGGACCTTGTCAATACCGCTCCAAATGCGACACAGTAT 1758
QY 380 ValGlyAsnAspPheSerTyrProTyrAspAsnAsnTyrAspGlyArgAsnValGlu 399
Db 1759 GTAGAAATGACTTACTTCGCCATACATGATTAATGGATGGCGCGCAATACGTAGAA 1818
QY 400 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 419
Db 1819 AATGATTTATTAATGACCAACAAAGCGGACGCTATACAAATGAAGTACAGGCTTAAAC 1878
QY 420 ValProSerGlyProGlnArgPheSerLeuAlaIleValHis 433
Db 1879 GTACCGTTGGACACAGAACTTCTCGTTGCAATTTGTGAAT 1920

RESULT 5
AAX37278
ID AAX37278 standard; DNA; 1923 BP.
XX AAX37278;
AC AAX37278;
XX 20-MAR-2003 (revised)
DT 21-JUL-1999 (first entry)
XX
DE Bacillus alkaline protease encoding DNA.

XX Alkaline protease; Bacillus; casein digestion; oleic acid; enzyme;
 KW washing composition; oxidising agent; ss.
 XX Bacillus sp.
 XX WO99182118-A1.
 XX 15-APR-1999.
 PD 07-OCT-1998; 98WO-JP004528.
 PF 07-OCT-1997; 97JP-00274570.
 XX (KAOS) KAO CORP.
 XX Takaiwa M, Okuda M, Saeki K, Kubota H, Hitomi J, Kageyama Y;
 PI Shikata S, Nomura M;
 XX WPI; 1999-287736/27.
 DR P-PSDB; AAY17088, AAY17090.
 XX Alkali protease from Bacillus used in washing powders.
 PT Disclosure; Page 58-63; 71pp; Japanese.
 XX The invention relates to alkaline proteases produced by strains of
 CC Bacillus. The proteases ability to digest casein is not inhibited by
 CC oleic acid and they have a high stability to oxidising agents. The
 CC alkaline protease of the invention has the following properties: (a) it
 CC is active over the pH range 4-13 and has at least 80% of its optimum
 CC activity over the range pH 6-12; (b) after 30 minutes at 40 deg. C it is
 CC stable over the pH range 6-11; c) its isoelectric point is 8.9-9.1; (d)
 CC its ability to digest casein is not inhibited by oleic acid; (e) it has
 CC molecular weight about 43,000 by SDS-PAGE. The alkaline proteases can be
 CC used as enzymes in washing compositions for use in automatic dishwashers
 CC and for washing clothes. The stability to oxidising agents allows the
 CC enzyme to be an effective component of washing compositions including
 CC bleaches. The present sequence represents an alkaline protease encoding
 CC DNA. (Updated on 20-MAR-2003 to correct DR field.)
 XX SQ Sequence 1923 BP; 576 A; 418 C; 475 G; 454 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 9,06e-149 Length: 1923
 Score: 1994.50 Matches: 380
 Percent Similarity: 94.01% Conservative: 28
 Best Local Similarity: 87.56% Mismatches: 25
 Query Match: 88.14% Indels: 1
 DB: 2 Gaps: 1

US-09-985-689a-3 (1-433) x AAX37278 (1-1923)

QY 1 AsnAspValAlaArgGlyIleValIysAlaAspValAlaGlnAsnAsnTyrGlyLeuTyr 20
 DB 619 AATGATGTTCGCGTGGAAATGTCTCAAGCGGATGTGCTCAGACGACTACGGGTGTAT 678

QY 21 GlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
 DB 679 GGACNAGGACAGATCGTAGCGGTTCGCATACAGGGCTTGATACAGTCCGATGACAGT 738

QY 41 SerMetHisGluAlaPheArgGlyIysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
 DB 739 TCGATGCATGAAGCCTTCGCGGGAAATTAATCTGATATATATGATTCGGACGCGCAAT 798

QY 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn--- 79
 DB 799 AATGCCAATGATACGAATGGTTCATGATCGCATGTGGCTGCTCCGTATTAGGAAACGGC 858

QY 80 AlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 99
 DB 859 TCCACTAATAAGGAATGCGGCTTCAGCGGAATCTAGTCTTCCAACTATCATCGATGATG 918

QY 100 SerGlyGlyLeuGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsn 119
 DB 919 GGTGGGGACTTGGAGGACTTACCTTCGAATCTCAAAACCTTATTACGCAAGCATACAGT 978

QY 120 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrAla 139
 DB 979 GCTGGTCCAGAATTCATACAAACTCTCTGGGGAGCAGTGAATGGGCTTACACAACA 1038

QY 140 AsnSerArgGlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAla 159
 DB 1039 GATTCAGAAATGTGGATGACTATGTGGCAAAATGATATGACGATCCTTTTCGTGTC 1098

QY 160 GlyAsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 179
 DB 1099 GGGAAATGAAGACCGAAGCGGGAACCATCAGTGCACAGCAGGATCAAAATGCAATA 1158

QY 180 ThrValGlyAlaThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsnProAsn 199
 DB 1159 ACAGTCGAGTACGGAACCTCCGCCCAAGCTTTGGGTCTTATGCGGACATATCAAC 1218

QY 200 HisIleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAsnVal 219
 DB 1219 CATGTGCACAGTTCTCTTCACGTGGACCGCAAGGATGCGGATCAAAACCGGATGTC 1278

QY 220 ThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 239
 DB 1279 ATGCAACCGGAACGTTTCATATATCAGCAAGATCTTCTTGCACCGGATTCCTCTTC 1338

QY 240 TrpAlaSerTyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 259
 DB 1339 TGGCGCAACCATGACAGTAAATATGATATGATGCTGGTGAACGTCCTCATCGGATC 1398

QY 260 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro 279
 DB 1399 GTTGTGGAACGTTGGACAGCTTCGTGAGCATTTCTGGAACACAGGACATCACACCA 1458

QY 280 LysProSerLeuIleLysAlaLeuAlaLeuIleAlaThrAspValGlyLeuGlyTyr 299
 DB 1459 AAGCTTCTCTATTAAAGCGGCACGTATTCGCGGTGCAGTGCATCGGCTTGGGTAC 1518

QY 300 ProSerGlyAspGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 319
 DB 1519 CCGAACGGTAACCAAGGATGGGACAGTGACATTTGATTAATCCCTGAACGTTGCCTAT 1578

QY 320 ValAsnGluAlaThrAlaLeuThrThrGlyGlnLysAlaThrTyrSerPheGlnThrGln 339
 DB 1579 GTGACAGAGTCCAGTCTCTATCCACAGCCCAAAAGCGACGTACTCGTTTACTGCTACT 1638

QY 340 AlaGlyLysProLeuLysIleSerLeuValThrThrAspAlaProGlySerThrAla 359
 DB 1639 GCCGCAAGCCTTTGAAATCTCCCTGTATGGTCTGATGCCCTTCGAGCACAACTGCT 1698

QY 360 SerTyrThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyGlnLysTyr 379
 DB 1699 TCCGTAACGTTGTCAATGATCTGGACCTTGTTCATTACCGCTCCAAATGGCACACATAT 1758

QY 380 ValGlyAsnAspPheSerTyrProTyrAspAsnAsnTrpAspGlyArgAsnValGlu 399
 DB 1759 GTAGGAATGACTTTACTTCGCCATACATATACCTGGATGGCGGCAATAACGTAGAA 1818

QY 400 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 419
 DB 1819 AATGATTTATTATTAATGCCCAAGCGGACGTATACAATAGGTACAGCTTATTAAC 1878

QY 420 ValProSerGlyProGlnArgPheSerLeuAlaIleValHis 433
 DB 1879 GTACCGGTTGGACCAACAGACCTTCTCGTTGGCAATTTGTAAT 1920

RESULT 6
 AAT85667
 ID AAT85667 standard; DNA; 1977 BP.
 XX
 AC AAT85667;

| | |
|--|---|
| XX | 17-OCT-2003 (revised) |
| DT | 20-APR-1998 (first entry) |
| XX | Thermococcus protease coding sequence. |
| DE | Protease; research reagent; thermal stability; thermococcus celer; ss. |
| KW | Thermococcus celer; DSM-2476. |
| XX | WO9721823-A1. |
| PN | 19-JUN-1997. |
| PD | 07-NOV-1996; 9SWO-JP003253. |
| Pf | 12-DEC-1995; 95JP-00323285. |
| XX | (TAKI) TAKARA SHUZO CO LTD. |
| PA | Takakura H, Morishita M, Yamamoto K, Mitta M, Asada K; |
| PI | Teunasaawa S, Kato I; |
| XX | WPI; 1997-332794/30. |
| DR | P-PSDB; AAW24121. |
| DR | Protease(s) and genes encoding them obtained from Thermococcus and |
| XX | Pyrococcus strains - have extremely high thermal stability and are useful |
| PT | industrially and as research reagents. |
| PT | Claim 3; Page 86-87; 159pp; Japanese. |
| XX | This sequence represents the coding sequence for the protease from |
| CC | Thermococcus celer DSM-2476. This sequence encodes a protease of the |
| CC | invention. The proteases of the invention have extremely high thermal |
| CC | stability. The proteases can be used as research reagents, and |
| CC | industrially in the food, drug and chemical industries. (Updated on 17- |
| CC | OCT-2003 to standardise OS field) |
| XX | Sequence 1977 BP; 453 A; 659 C; 554 G; 311 T; 0 U; 0 Other; |
| SQ | |
| Alignment Scores: | |
| Pred. No.: | 1,09E-25 Length: 1977 |
| Score: | 443.00 Matches: 137 |
| Percent Similarity: | 46.03% Conservative: 66 |
| Best Local Similarity: | 31.07% Mismatches: 170 |
| Query Match: | 19.58% Indels: 68 |
| DB: | Gaps: 16 |
| US-09-985-689A-3 (1-433) x AAT85667 (1-1977) | |
| Qy | 8 VallyeAlaaspValalaglnAsnAsntyrrGlyLeutyrrGlyGlnGlnValValala 27 |
| Db | 433 ATAGGGCCCGATACCGTCGTGGAACTCCCTCGCTACGACGAGCGGTGTGGTTGCC 492 |
| Qy | 28 ValAlaaspThrClyLeuaspThrGlyArgasnAspSerMetHisGlualaPheArg 47 |
| Db | 493 ATCGTCGATACGGGTATAGACCGAAC-----CACCCCGATCTCGAG 534 |
| Qy | 48 GlyLysallethrAlalaLeutyrr--AlalaLeuGlyArgThrAsnAsnAlaAsnAspProasn 66 |
| Db | 535 GGCAAGTCATAGGTCGTACTACACCGCGTCAACGCAGGTGCGACCCCTACGATGCCAG 594 |
| Qy | 67 GlyHisGlyThrHisValaladGlySerValleuGly-----AsnAlaLeuAsn 82 |
| Db | 595 GGACACGGAAACCACCGTTTCGGGTATCGTTGCCGAACCGGCAGCGGTAACTCCCAGTAC 654 |
| Qy | 83 LysClyMetalaproGlnAlaAsnLeuValPheGlnSerlleMet-----AspSerSer 100 |
| Db | 655 ATAGCGTGTCGCCCGCGCGAAGTCGTGCGCGTCAAGGTTCCTCGGTGCCGACGGTTCG 714 |
| Qy | 101 GlyGlyLeuGlyGlyLeuProserAsnLeuAsnThrLeuPheSerGlnAlaTrpAsnAla 120 |

Db 715 GGAAGCGTCTCCACCATCATCGGGGTGTTGACTGGTGCTCCAGACAAGCAGACAGTAC 774
 Qy 121 GlyAlaargileHistrAsnSerTrpIglyValaProValasnGlyAlaTyThrAlaAsn 140
 Db 775 GGGATAAGGGTCATCAACCTTCCCTCGGCTCCTCCAGAGCTCCGACGAAACCGACTCC 834
 Qy 141 SerArgGlnValAspGluTyValArgAsnAsnAspMetThrValLeuPheAlaAagly 160
 Db 835 CTCAGTCAGGCCCTCAACAAGCGCTGGAGCGCGGTATAGTAGTCTCGCTCGCCGCCGC 894
 Qy 161 AsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlatysAsnAlaIleThr 180
 Db 895 AACAGCGGGCGGAACACTACACGTCGGCTCACCGCGCGCGAGCAAGGTCATAACC 954
 Qy 181 ValGlyAlaThrGluAsnTyArgProSerPheGlySerIleAlaAspAsnProAsnHis 200
 Db 955 GTCCGTGCA-----GTTGACACCAACGACAC 981
 Qy 201 IleAlaGlnPheSerArgGlyAlaThrArgAspGlyArgIleLysProAspValThr 220
 Db 982 ATGCCAGCTTCTCCAGACGGGACCGACCGCGGAGAGGCTCAAGCGGAGTGTCTC 1041
 Qy 221 AlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrp 240
 Db 1042 GCCTCCGGCGCTTGACATCATAGCCCGCGCGCCAGC-----GGAACCGACGATCGGC 1092
 Qy 241 AlaAsnTyArgSerLysTyAlaTyMetGlyGlyThrSerMetAlaThrProIleVal 260
 Db 1093 ACCCGGTAACGACTACTACCAAGGCCTCTGGAACCGAGATGGCCACCCCGCAGTT 1152
 Qy 261 AlaGlyAsnValAlaGlnLeuA^agIUHPheIleLysAsnArgGlyIleThrPro--- 279
 Db 1153 TCGGGCGCTTGGCGCGCTCATCTCCAG-----GCCACCGAGCTGGACCCCGGAC 1203
 Qy 280 -----LysProSerIleLysAlaAlaLeuIleAlaGlyAlaThrAspValGly-- 296
 Db 1204 AAGGTGAAGACCGCCCTCATCGAGACCGCCGACATAGTCGCCCCCAAGGAGATAGCGGAC 1263
 Qy 297 LeuGlyTyPrProSerGlyAspGlnGlyTrpIglyArgValThrLeuAspLysSerLeuAsn 316
 Db 1264 ATGCCTTAC-----GGTGGGTAGGTGAACGCTACAAGGCC-----Lys 331
 Qy 317 ValAlaTyValAsnGluAlaThrAlaLeuThrThrdGlyGln-----Lys 331
 Db 1303 ATCAAGTACGACGACTACGCAAGCTCCTTCACCGGCTCGGTGCGCGCACAGGGAAGC 1362
 Qy 332 AlaThrTySerPheGlnThrGlnAlaGlyLysProLeuLysIleSerLeuValTrpThr 351
 Db 1363 GCCACCCACACCTTCGAGCTCAGCGGCGCCACCTTCGTGACCGGCCACCTCTACTGGGAC 1422
 Qy 352 AspAlaProGlySerThrAlaSerTyThrLeuValAsnAspLeuAspLeuValIle 371
 Db 1423 -----ACGGGCTCGAGCGACATCGACCTCTACCTC 1452
 Qy 372 ThrAlaProAsnGlyGlnLysTyValIglyAsnAspPheSerTyPrProTyArgAsnAsn 391
 Db 1453 TAGACCCCCAAC-----GGAGAAGAGTTGACTACTCTCTACACCGCCCTAC 1497
 Qy 392 TrpAspGlyArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThrTy 411
 Db 1498 TAC-----GGCTTCGAAGAGTCCGCTACTACAACCGACCGCGGAACCTGG 1545
 Qy 412 ThreIleGluValGlnAlaTyArgAsnValProSerGlyProGlnArgPheSerLeuAlaIle 431
 Db 1546 ACGGTCAAGGTGTCAGCTAC-----AAGGGCGCGGGAACCTACAGGTGCAGCTC 1596
 Qy 432 Val 432
 Db 1597 GTC 1599

RESULT 7
AAX05926
ID AAX


```

QY 432 Val 432
Db 1597 GTC 1599

RESULT 8
AAK05920
ID AAX05920 standard; DNA; 1236 BP.
XX
AC AAX05920;
XX
DT 06-MAY-1999 (first entry)
XX
DE Hyperthermostable protease fragment encoding DNA.
XX
KW Hyperthermostable; protease; thermophilic; bacterium; subtilisin;
KW additive; drug; washing agent; foodstuff; chemical synthesis; ds.
XX
OS Pyrococcus furiosus.
XX
PN WO9856926-A1.
XX
PD 17-DEC-1998.
XX
XX 04-JUN-1998; 98WO-JP002465.
XX
XX 10-JUN-1997; 97JP-00151969.
XX
XX (TAKI ) TAKARA SHUZO CO LTD.
XX
XX Takakura H, Morishita M, Shimojo T, Asada K, Kato I;
XX
XX WPI; 1999-080907/07.
XX
XX P-PSDB; AAW94836.
XX
XX Recombinant hyperthermostable protease from Pyrococcus furiosus - and
XX gene encoding it, for large scale production of the protease for
XX industrial use.
XX
XX Claim 6; Page 37-38; 82pp; Japanese.
XX
XX The invention relates to a hyperthermostable protease derived from a
XX thermophilic bacterium (especially Pyrococcus furiosus). The protease has
XX working temperature 40-110 deg.C (optimum 80-95 deg.C), working pH 5-10
XX (optimum 6-8), and retains more than 90% of its activity after 8 hours at
XX 95 deg.C. The invention also provides gene sequences encoding a
XX polypeptide of formula SIG-Ala-Gly-Ans-PRO, where SIG is a signal
XX peptide from subtilisin, and PRO is the above protease. Host cells
XX (especially Bacillus strains) transformed with vectors comprising the
XX genes are used for the recombinant production of the protease. The
XX hyperthermostable protease which can be prepared in quantity suitable for
XX industrial use, can be used as an additive for drugs, washing agents and
XX foodstuffs and for chemical synthesis
XX
XX Sequence 1236 BP; 368 A; 251 C; 309 G; 308 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 2,6e-24 Length: 1236
Score: 422.50 Matches: 138
Percent Similarity: 43.96% Conservative: 55
Best Local Similarity: 31.44% Mismatches: 167
Query Match: 18.67% Indels: 79
DB: 2 Gaps: 17

US-09-985-689A-3 (1-433) x AAX05920 (1-1236)

QY 16 AsnTyrGlyLeuTyrGlyGlnValValAlaValAlaAspThrGlyLeuAspThr 35
Db 58 AACTGGGATATGATGCTCTGGATATCAATAGGAATTAATGACCTGGAATTGAC--- 114
QY 36 GlyArgAsnAspSerSerMetHisGluAlaPheArgGlyValLeuThrAlaLeuTyrAla 55
Db 115 -----GCTTCTCATCCAGATCTCCAGGAAGAAGTA----- 144

```

Db 1096 GCTACTATGATTCGAAAGGTTGGTTATTACAAACCACTGATGGACATGGCAATT 1155
 QY 414 GluValGlnAlaTyrAsnValProSerGlyProGlnArgPheSerLeuAlaIleVal 432
 Db 1156 AAGTTGTAGCTAC-----ACGGAAGTGCAAACTATCAAGTAGATGTGTA 1203

RESULT 9

AAT85668 standard; DNA; 1566 BP.
 AC AAT85668;

DT 17-OCT-2003 (revised)
 DT 20-APR-1998 (first entry)

XX Pyrococcus furiosus protease coding sequence.

KW Protease; research reagent; thermal stability; pyrococcus furiosus; ss.

XX Pyrococcus furiosus; DSM-3638.

PH Key Location/Qualifiers
 CDS 1..1566

FT /tag= a

FT /transl_except= (pos: 1282..1284, aa: Xaa)

FT /note= "Xaa= Gly, Val"

XX WO9721823-A1.

PD 19-JUN-1997.

XX 07-NOV-1996; 96WO-JP003253.

XX 12-DEC-1995; 95JP-00323285.

XX (TAKI) TAKARA SHUZO CO LTD.

XX Takakura H, Morishita M, Yamamoto K, Mitta M, Asada K;

PI Teunasawa S, Kato I;

XX WPI; 1997-332794/30.

DR P-PSDB; AAW24122.

XX Protease(s) and genes encoding them obtained from Thermococcus and
 PT Pyrococcus strains - have extremely high thermal stability and are useful
 PT industrially and as research reagents.

XX Claim 7; Page 90-91; 159pp; Japanese.

CC This sequence represents the coding sequence for the protease from
 CC Pyrococcus furiosus DSM-3638. This sequence encodes a protease of the
 CC invention. The proteases of the invention have extremely high thermal
 CC stability. The proteases can be used as research reagents, and
 CC industrially in the food, drug and chemical industries. (Updated on 17-
 CC OCT-2003 to standardise OS field)

XX SQ Sequence 1566 BP; 467 A; 342 C; 372 G; 384 T; 0 U; 1 Other;

Alignment Scores:

Pred. No.: 3,47e-24 Length: 1566
 Score: 422.50 Matches: 138
 Percent Similarity: 43.96% Conservative: 55
 Best Local Similarity: 31.44% Mismatches: 167
 Query Match: 18.67% Indels: 79
 DB: 2 Gaps: 17

US-09-985-689A-3 (1-433) x AAT85668 (1-1566)

QY 16 AsnTyrGlyLeuTyrGlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThr 35

Db 58 AACTGGGATGATGATGTTCTGGATCACAATAGGAATATTGACACTGGATTGAC--- 114

QY 36 GlyArgAsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAla 55
 Db 115 -----GCTTCTCATCCAGATCTCCAGGAAAGTA----- 144
 QY 56 LeuGlyArgThrAsnAsnAlaAsn-----AspProAsnGlyHisGly 69
 Db 145 ATTGGGTGGGTAGATTGTCATGTTAGGAGTTATCCATACGATGACCATGGACATGGA 204
 QY 70 ThrHisValAlaGlySerValLeuGly-----AsnAlaLeuAsnLysGly 84
 Db 205 ACTCATGTAGCTCAATAGCAGCTGGTACTGGAGCAGCAAGTAATGGCAAGTCAAGGGA 264
 QY 85 MetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerSerGlyGly 102
 Db 265 ATGGCTCCAGGAGCTAAGCTGGCGGGAATTAAAGTTCTAGTCCGATGGTTCTGGAGC 324
 QY 103 LeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsnAlaGlyAla 122
 Db 325 ATATCTACTATAATTAAGGAGTTGAGTGGCGCGTTGATTAACAAGATAGTAGCGAATT 384
 QY 123 ArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrAlaAsnSerArg 142
 Db 385 AAGTCATTAATCTTCTCTTGGTTCAAGCCAGAGCTCAGATGGTACTGACGCTCTAAGT 444
 QY 143 GlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAlaGlyAsnGlu 162
 Db 445 CAGGCTGTTAATGAGCGCTGGGATCTGGATTAGTTGTTGTTGCTGGCTGGAACAGT 504
 QY 163 GlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGly 182
 Db 505 GGACCTAACAAAGTATCAATCGGTTCTCCAGAGCTGCAAGCAAAAGTTATTACATTGGA 564
 QY 183 AlaThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsnProAsnIleAla 202
 Db 565 GCCGTTGACAAGTAT-----GATGTTATAACA 591
 QY 203 GlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspValThrAlaPro 222
 Db 592 AGCTTCTCAAGCAGAGGGCAACTGCAGAGCGAGGCTTAAGCCTGAGGTGTGTGTCCA 651
 QY 223 GlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsn 242
 Db 652 GGAAACTGGATAATTGCTGCCAGAGCAAGT-----GGAAC TAGCATGGGTCAACCA 702
 QY 243 TyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGly 262
 Db 703 ATTAATGACTATTACACAGCAGCTCTCTGGACATCAATGGCAACTCTCTCAGTAGTGT 762
 QY 263 AsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro----- 279
 Db 763 ATTGCAGCCCTCTTGCTCCAA-----GCACACCCGAGCTGGACTCCAGACAAAGTA 813
 QY 280 LysProSerLeuIleLysAlaAlaLeuIleAlaGlyAlaThrAspValGly----LeuGly 298
 Db 814 AAAACAGCCCTCATAGAAACTGCTGATATCGTAAGCCAGATGAATAGCGATATATGCC 873
 QY 299 TyrProSerGlyAspGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAla 318
 Db 874 TAC-----GGTCAGGTAGGTTAATGCATACAGCGGTATAAAC----- 912
 QY 319 TyrValAsnGluAlaThrAlaLeuThrThrGlyGlnLysAla-----Thr 333
 Db 913 TACGATAACTATGCAAAAGCTAGTGTTCACCTGGATATGTCGCAACAAAGGCGAGCCAACT 972
 QY 334 TyrSerPheGlnThrGlnAlaGlyLysProLeuLysIleSerLeuValTrpThrAspAla 353
 Db 973 CACCAGTTCGTTATTAGCGAGCTTCGTTGTAACCTCCACATTAATCTGGACAATGCC 1032
 QY 354 ProGlySerThrThrAlaSerTyrThrLeuValAsnAspLeuAspLeuValIleThrAla 373
 Db 1033 AAT-----AGCGACCTTGATCTTTACCTCTACCAT 1062
 QY 374 ProAsnGlyGlnLysTyrValGlyAsnAspPheSerTyrProTyrAspAsnAsnTrpAsp 393

```

Db      1063 CCCATGGAACACAGGTT-----GACTACTCTTAC-----ACC 1095
Qy      394 GlyArgAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThrThrIle 413
Db      1096 GCCTACTATGATTCGAAAAGGTGGTTATTACAAACCACTGATGGACATGACAAATT 1155
Qy      414 GluValGlnAlaTyrAsnValProSerGlyProGlnArgPheSerLeuAlaIleVal 432
Db      1156 AAGGTTGTAAGCTAC-----AGCGAAGTGCAAACTATCAAGTAGATGTGGTA 1203

RESULT 10
ID      AAT85695
XX      AAT85695 standard; DNA; 1962 BP.
XX      AAT85695;
XX
XX      17-OCT-2003 (revised)
DT      20-APR-1998 (first entry)
XX
XX      Pyrococcus furiosus PFUS protease coding sequence.
XX
XX      Protease; research reagent; thermal stability; pyrococcus furiosus; ss.
XX
XX      Pyrococcus furiosus; DSM-3638.
XX      WO9721823-A1.
XX      19-JUN-1997.
XX
XX      07-NOV-1996; 96WO-JP003253.
XX
XX      12-DEC-1995; 95JP-00323285.
XX      (TAXI ) TAKARA SHUZO CO LTD.
XX
XX      Takakura H, Morishita M, Yamamoto K, Mitta M, Asada K;
PI      Tsunawasa S, Kato I;
XX
XX      WPI; 1997-332794/30.
DR      P-PSDB; AAM24129.
XX
XX      Protease(s) and genes encoding them obtained from Thermococcus and
PT      Pyrococcus strains - have extremely high thermal stability and are useful
PT      industrially and as research reagents.
XX
XX      Disclosure; Page 123-125; 159pp; Japanese.
XX
XX      This sequence represents the coding sequence for the protease from
CC      Pyrococcus furiosus DSM-3638. This sequence encodes a protease of the
CC      invention. The proteases of the invention have extremely high thermal
CC      stability. The proteases can be used as research reagents, and
CC      industrially in the food, drug and chemical industries. (Updated on 17-
CC      OCT-2003 to standardise OS field)
XX
XX      Sequence 1962 BP; 602 A; 399 C; 471 G; 490 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:      4,569-24      Length:      1962
Score:          422.50      Matches:      138
Percent Similarity: 43.96%      Conservative: 55
Best Local Similarity: 31.44%      Mismatches: 167
Query Match:     18.67%      Indels:      79
DB:             2          Gaps:      17

US-09-985-689A-3 (1-433) x AAT85695 (1-1962)

Qy      16 AsnTyrGlyLeuTyrGlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThr 35
Db      454 AACTTGGGATATGATGGTTCTCGAATCAATAGGAATAATGACACTGGAATTGAC--- 510
Qy      36 GlyArgAsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAla 55

```

```

511 -----GCTTCTCATCAGATCTCCAGGAAAGTA----- 540
Qy      56 LeuGlyArgThrAsnAsnAlaAsn-----AspProAsnGlyHisGly 69
Db      541 ATTGGTGGGTAGATTTTGTCAATGATAGGAGTTATCCATACGATCACCATGCATGGA 600
Qy      70 ThrHisValAlaGlySerValLeuGly-----AsnAlaLeuAsnLysGly 84
Db      601 ACTCATGTAGCTTCAATAGCAGCTGTACTGAGCAGCAGCAAGTAATGGCAAGTACAGGA 660
Qy      85 MetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerSerGlyGly 102
Db      661 ATGGCTCCAGGAGCTAAGCTGGCGGGAATTAAAGTTTCTAGGTGCCGATGGTCTTGAAGC 720
Qy      103 LeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTyrAsnAlaGlyAla 122
Db      721 ATATCTACTATAATTAAGGAGTGTAGTGGCGGTGATTAACAAGATAAAGTACGGAATT 780
Qy      123 ArgIleHisThrAsnSerTyrGlyAlaProValAsnGlyAlaTyrThrAlaAsnSerArg 142
Db      781 AAGGTCAATTAATCTTTCTCTGTTCAAGCCAGAGCTCAGATGCTGACTGACGCTCTAAGT 840
Qy      143 GlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAlaGlyAsnGlu 162
Db      841 CAGGCTGTTAATGCACGCGTGGATGCTGATTAGTTGTGTGTTGCCCGCTGGAACAGT 900
Qy      163 GlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGly 182
Db      901 GGACCTAAACAAGTATACATCGTTCTCCAGCAGCTGCAGCAAGATTATTACAGTTGA 960
Qy      183 AlaThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsnProAsnHisIleAla 202
Db      961 GCGTTTGACAAGTAT-----GATGTTATAACA 987
Qy      203 GlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspValThrAlaPro 222
Db      988 AGCTTCTCAAGCAGAGGCGCACTGCAGACGCGAGCTTAAGCCTGAGGTTTGTCTCCA 1047
Qy      223 GlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsn 242
Db      1048 GGAAACTGGATAATTGCTGCCAGAGCAAGT-----GGAACTAGCATGGTGTCAACCA 1098
Qy      243 TyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGly 262
Db      1099 ATTAATGACTATTACACAGCAGCTCTGGGACATCAATGGCACTCCTCAGTAGTGTGT 1158
Qy      263 AsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro----- 279
Db      1159 ATTGCAGCCCTCTTGTCTCAA-----GCACACCCCGAGCTGGACTCCACACAAAGTA 1209
Qy      280 LysProSerLeuIleLysAlaLeuIleAlaGlyAlaThrAspValGly---LeuGly 298
Db      1210 AAAACAGCCCTCATAGAAACTGCTGATATCGTAAAGCCAGATGAATAGCCGATATAGCC 1269
Qy      299 TyrProSerGlyAspGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAla 318
Db      1270 TAC-----GGTGCAGTAGGTTTAATGCATACAGGCTATAAAC----- 1308
Qy      319 TyrValAsnGluAlaThrAlaLeuThrThrGlyGlnLysAla-----Thr 333
Db      1309 TACGATAACTATGCAAAAGTAGTGTTCAGTGTATGTGGCCAAACAAAGGAGCCAAACT 1368
Qy      334 TyrSerPheGlnThrGlnAlaGlyLysProLeuLysIleSerLeuValTyrThrAspAla 353
Db      1369 CACCAGTTGTTATTAGCGGAGCTTCGTTCGTAACTGCCACATTATCTGGGCAATGCC 1428
Qy      354 ProGlySerThrThrAlaSerTyrThrLeuValAsnAspLeuAspLeuValIleThrAla 373
Db      1429 AAT-----AGCGACCTTGATCTTTTACCTCTACAGT 1458
Qy      374 ProAsnGlyGlnLysTyrValGlyAsnAspPheSerTyrProTyrAspAsnAsnTrpAsp 393
Db      1459 CCCAATGGAAACACAGGTT-----GACTACTCTTAC-----ACC 1491

```

QY 394 GlyArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThrThrThrThr 413
 Db 1492 GCTACTATGATTCGAAAGGTGGTTATTACACCCCACTGATGGACATGGACAT 1551
 QY 414 GluValGlnAlaTyrAsnValProSerGlyProGlnArgPheSerLeuAlaIleVal 432
 Db 1552 AAGTTTGAAGCTAC-----AGCGAAGTGCAAACTATCAAGTAGATGTGTA 1599

RESULT 11

AA05929
 ID AAX05929 standard; DNA; 1962 BP.

XX AC AAX05929;

XX DT 06-MAY-1999 (first entry)

XX Hyperthermostable protease encoding DNA.

XX Hyperthermostable; protease; thermophilic; bacterium; subtilisin;
 KW additive; drug; washing agent; foodstuff; chemical synthesis; ds.

XX Pyrococcus furiosus.

XX PN WO9856926-A1.

XX PD 17-DEC-1998.

XX PF 04-JUN-1998; 98WO-JP002465.

XX PR 10-JUN-1997; 97JP-00151969.

XX PA (TAKI) TAKARA SHUZO CO LTD.

XX PI Takakura H, Morishita M, Shimojo T, Asada K, Kato I;

XX DR WPI; 1999-080907/07.

XX DR P-PSDB; AAW94841.

XX Recombinant hyperthermostable protease from Pyrococcus furiosus - and
 PT gene encoding it, for large scale production of the protease for
 PT industrial use.

XX PS Disclosure; Page 59-60; 82pp; Japanese.

XX The invention relates to a hyperthermostable protease derived from a
 CC thermophilic bacterium (especially Pyrococcus furiosus). The protease has
 CC working temperature 40-110 deg.C (optimum 80-95 deg.C), working pH 5-10
 CC (optimum 6-8), and retains more than 90% of its activity after 8 hours at
 CC 95 deg.C. The invention also provides gene sequences encoding a
 CC polypeptide of formula SIG-Ala-Gly-Asn-PRO, where SIG is a signal
 CC peptide from subtilisin, and PRO is the above protease. Host cells
 CC (especially Bacillus strains) transformed with vectors comprising the
 CC genes are used for the recombinant production of the protease. The
 CC hyperthermostable protease which can be prepared in quantity suitable for
 CC industrial use, can be used as an additive for drugs, washing agents and
 CC foodstuffs and for chemical synthesis

XX SQ Sequence 1962 BP; 602 A; 399 C; 471 G; 490 T; 0 U; 0 Other;

Alignment Scores:

| | | | |
|------------------------|----------|---------------|------|
| Pred. No.: | 4,56e-24 | Length: | 1962 |
| Score: | 422.50 | Matches: | 138 |
| Percent Similarity: | 43.96% | Conservative: | 55 |
| Best Local Similarity: | 31.44% | Mismatches: | 167 |
| Query Match: | 18.67% | Indels: | 79 |
| DB: | 2 | Gaps: | 17 |

US-09-985-689A-3 (1-433) x AAX05929 (1-1962)

QY 16 AsnTyrGlyLeuTyrGlyGlnValValAlaValAlaAspThrGlyLeuAspThr 35
 Db 454 AACTTGGGATGATGATGTTCTGGATCACAATAGGAATATTGACATGGAAATTGAC--- 510

QY 36 GlyArgAsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAla 55
 Db 511 -----GCTTCTCATCCAGATCTCCAAAGGAAAGTA----- 540
 QY 56 LeuGlyArgThrAsnAsnAlaAsn-----AspProAsnGlyHisGly 69
 Db 541 ATTGGGTGGGTAGATTTTGTCAATGGTAGAGAGTTATCCATAGATGACCATGACATGGA 600
 QY 70 ThrHisValAlaGlySerValLeuGly-----AsnAlaLeuAsnLysGly 84
 Db 601 ACTCATGTAGCTTCAATAGCAGCTGTTACTGGAGCAGCAAGTAATGGCAAGTACAAGGGA 660
 QY 85 MetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerSerGlyGly 102
 Db 661 ATGCTCCAGGAGCTAAGCTGGCGGAATTAAGTTCTTAGTGCCGATGTTCTTGAAGC 720
 QY 103 LeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsnAlaGlyAla 122
 Db 721 ATATCTACTATAATTAAGGGAGTTAGTGGCGGTTGATAACAAAGATAAGTACGGAAT 780
 QY 123 ArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrAlaAsnSerArg 142
 Db 781 AAGTCAATTAATCTTCTCTGTTTCAAGCCAGAGCTCAGATGGTACTGACGGCTCTAAGT 840
 QY 143 GlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAlaGlyAsnGlu 162
 Db 841 CAGGCTGTTAATGCAGCGTGGGATGCTGGATTAGTTGTTGTTGCTGCCGCTGAAACAGT 900
 QY 163 GlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGly 182
 Db 901 GGACCTAACCAAGTATACATCGGTTCTCCAGCAGCTGCAAGCAAGATTATTACAGTTGA 960
 QY 183 AlaThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsnProAsnHisIleAla 202
 Db 961 GCCGTTGCAAGTAT-----GATGTTATAACA 987
 QY 203 GlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspValThrAlaPro 222
 Db 988 AGCTTCTCAAGCAGAGGCCCAACTGCAGACGGCAGGCTTAAGCCTGAGTTGTGTCTCA 1047
 QY 223 GlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsn 242
 Db 1048 GGAACCTGGATTAATTGCTGCCAGAGCAAGT-----GGAACCTAGCATGGGTCAACCA 1098
 QY 243 TyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGly 262
 Db 1099 ATTAATCACTATTACACAGCAGCTCTCGGACATCAATGGCAACTCTCTCACGTAGCTGT 1158
 QY 263 AsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro----- 279
 Db 1159 ATTGCAGCCCTCTTGTCTCCAA-----GCACCCCGAGCTGGACTCCAGACAAAGTA 1209
 QY 280 LysProSerLeuIleLysAlaAlaLeuIleAlaGlyAlaThrAspValGly---LeuGly 298
 Db 1210 AAACAGCCCTCATAGAACTGCTGATATCGTAAAGCCAGATGAATAAGCCGATATAGCC 1269
 QY 299 TyrProSerGlyAspGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAla 318
 Db 1270 TAC-----GGTGCAGGTAGGTTAATGCTATACAGGCTATAAAC----- 1308
 QY 319 TyrValAsnGluAlaThrAlaLeuThrThrGlyGlnLysAla-----Thr 333
 Db 1309 TACGATAACTATGCAAGCTAGTCTTCACTGGATATGTTGCCAACAAAGGCAGGCCAACT 1368
 QY 334 TyrSerPheGlnThrGlnAlaGlyLysProLeuLysIleSerLeuValThrAspAla 353
 Db 1369 CACCAGTTTCGTTATTAGCGGAGCTTCGTTCTGTAACCTGCCACATTATATGGAATGCCC 1428
 QY 354 ProGlySerThrThrAlaSerTyrThrLeuValAsnAspLeuAspLeuValIleThrAla 373
 Db 1429 AAT-----AGCAGCTTGTATCTTTACCTCTACGAT 1458

```

Qy 374 ProAenGlyGlnLysTyrValGlyAsnAspPheSerTyrProTyrAspAsnAsnTrpAsp 393
Db 1459 CCCAATGGAAACACAGGTT-----GACTACTCTTAC-----ACC 1491
Qy 394 GlyArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIle 413
Db 1492 GCCTACTATGATTGCAAGAGTTTGGTTATTACACCCAACTGATGGAACATGGACAATT 1551
Qy 414 GluValGlnAlaTyrAsnValProSerGlyProGlnArgPheSerLeuAlaIleVal 432
Db 1552 AAGGTTGTAAGCTAC-----AGCGGAAGTCAAACTATCAAGTAGATGTGGTA 1599

RESULT 12
AAT85669
ID AAT85669 standard; DNA; 1977 BP.
XX
AC AAT85669;
XX
DT 20-APR-1998 (first entry)
XX
DE Protease coding sequence.
XX
KW Protease; research reagent; thermal stability; ss.
XX
OS Synthetic.
XX
PN WO9721823-A1.
XX
PD 19-JUN-1997.
XX
XX 07-NOV-1996; 96WO-JP003253.
XX
XX 12-DEC-1995; 95JP-00323285.
XX
PR (TAKI ) TAKARA SHUZO CO LTD.
PA Takakura H, Morishita M, Yamamoto K, Mitta M, Asada K;
PI Tsunasawa S, Kato I;
XX
WPI: 1997-332794/30.
DR P-PSDB; AAW24123.
DR
XX
PT Protease(s) and genes encoding them obtained from Thermococcus and
PT Pyrococcus strains - have extremely high thermal stability and are useful
PT industrially and as research reagents.
XX
PS Claim 11; Page 95-97; 159pp; Japanese.
XX
SQ Sequence 1977 BP; 527 A; 562 C; 513 G; 375 T; 0 U; 0 Other;

Alignment Scores:
Pred. NO.: 2,13e-22 Length: 1977
Score: 401.50 Matches: 130
Percent Similarity: 43.96% Conservative: 63
Best Local Similarity: 29.61% Mismatches: 167
Query Match: 17.74% Indels: 79
DB: 2 Gaps: 17

US-09-985-689A-3 (1-433) x AAT85669 (1-1977)

Qy 16 AsnTyrGlyLeuTyrGlyGlnGlyValValAlaValAlaAspThrGlyLeuAspThr 35
Db 454 AACTTGGATATGATGGTCTTGGATCATCATAGGAATAATTGACACTGGATTGAC--- 510
Qy 36 GlyArgAsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAla 55
Db 511 -----GCTTCTCATCGATCTCCAAGGAAAAGTA----- 540

```

Db 1501 -----GGCTTCGAGNAGTCGGCTACTACACCGCGGCGGACCTGACGGTC 1551
Qy 414 GluValGlnAlaTyrAsnValProSerGlyProGlnArgPheSerLeuAlaIleVal 432
Db 1552 AAGGTCGTCAGCTAC-----AAGGCGCGCGGCGGAACTACAGGTCGACGTCGTC 1599
RESULT 13
ID ABL54900 standard; DNA; 2121 BP.
XX ABL54900;
AC XX
XX
DT 11-SEP-2003 (revised)
DT 31-MAY-2002 (first entry)
XX
XX
DE T. yonsei subtilisin-like serine protease coding sequence.
XX
XX Subtilisin-like serine protease; ss.
KW
XX
OS Thermoanaerobacter yonseiensis.
XX
XX
FH Key Location/Qualifiers
FT 142..1179
FT /*tag= a
FT /product= "subtilisin-like serine protease"
XX
XX KR2000072141-A.
XX
XX 05-DEC-2000.
XX
XX 04-AUG-2000; 2000KR-00045411.
XX
XX 04-AUG-2000; 2000KR-00045411.
XX (KIMV/) KIM Y S.
XX
XX Chang HJ, Kim DH, Byun YR, Kim YS;
XX WPI; 2001-298092/31.
XX P-PSDB; ABB09483.
XX
XX New DNA sequence of thermophilic protein decomposition enzyme and protein
XX derived therefrom.
XX
XX Claim 1; Page 6; 15pp; Korean.
XX
XX This sequence represents the DNA encoding the Thermoanaerobacter yonsei
XX subtilisin-like serine protease of the invention. (Updated on 11-SEP-2003
XX to standardise OS field)
XX
XX Sequence 2121 BP; 712 A; 410 C; 425 G; 554 T; 0 U; 20 Other;
SQ

Alignment Scores:
Pred. No.: 6.68e-20 Length: 2121
Score: 370.50 Matches: 131
Percent Similarity: 45.85% Conservative: 57
Best Local Similarity: 31.95% Mismatches: 139
Query Match: 16.37% Indels: 83
DB: 4 Gaps: 18

US-09-985-689a-3 (1-433) x ABL54900 (1-2121)
Qy 6 GlyIleValIysAlaAspValAlaGlnAsnTyrGlyLeuTyrGlyGlnGlyGlnVal 25
Db 445 GGAATCACAAA-----GCACGAGTGAATTTGGAGTCACAGGAAAAAATAAACA 495
Qy 26 ValIleValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerMetHisGluAla 45
Db 496 ATACGAATATTGACACAGGTATAGCGGAATCACGTTCACTCTCTCA----- 543
Qy 46 PheArgGlyIleIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAsn----- 63

Db 544 ---GGTGGAAAAATA-----ATAGATGGAAAGACTTTTATCAACAACAA 585
Qy 64 -----AspProAsnGlyHisGlyThrHisVal-----AlaGlySer 75
Db 586 ACTACACCATACGACGACATGGCCATGGAACTCAGGTAGCAAGTATTCTCGAGGTACA 645
Qy 76 ValLeuGlyAsnAlaLeuAsnIleGlyMetAlaProGlnAlaAsnLeuValPheGlnSer 95
Db 646 GGTGCTGGAAACAGTCTTTACAAAGCGTGTCTCTGATGCTTTGTTGTAGTAATAAA 705
Qy 96 IleMetAspSerSerGly-----GlyLeuGlyGlyLeu 106
Db 706 GTTTTAGATGCAATGGAAGCGGCACATGAGCAGCTGTAACCTGCAGGAATTGACTGGCT 765
Qy 107 ProSerAsnLeuAsnThrLeuPheSerGlnAlaTyrAsnAlaGlyAlaArgIleHisThr 126
Db 766 GTTCAAAATAAAGATGTATACGGAATCAAAAGTTATAAATTTAAGCTCGCCACTTCTACA 925
Qy 127 AsnSerTrpGlyAlaProValaAsnGlyAlaTyrThrAlaAsnSerArgGlnValAspGlu 146
Db 826 AGTCTGATGGA-----ACTGACTCTCTACTCTATTAGCAGTGAATAGACAGTAGAT--- 876
Qy 147 TyrValArgAsnAsnAspMetThrValLeuPheAlaAlaGlyAsnGluGlyProAsnSer 166
Db 877 -----AGCGGTATTGTAGTGTGTAGCAGCAGGAAACTCTGCGCCCTGCAAAA 924
Qy 167 GlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsn 186
Db 925 TACACCATAGGTCCCTGCTGCTGGGAAAGAACCTACATTCACGTCGACGATGGCAGAT 984
Qy 187 TyrArgProSerPheGlySerIleAlaAspAsnProAsnHisIleAlaGlnPheSerSer 206
Db 985 -----GTAGTGAACCTTGGCTTTAAC-----CTTCAAGCTTTTCACG 1023
Qy 207 ArgGlyAlaThrArgAspGlyArgIleLysProAspValThrAlaProGlyThrPheIle 226
Db 1024 CGCGTCTTACTGCTGACGGAAGATAAAACCTGACATTCGCGCCCGCAGGATATAATATA 1083
Qy 227 LeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsnTyrAsnSerLys 246
Db 1084 ACTGCGCGCAAG-----CGCAATTCGTAAATGGA 1113
Qy 247 TyrAlaTyrMetGlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGln 266
Db 1114 TATGTAACATACACGCGGTACAGCATGGCAACACCTTTGTAGCAGGAACCTGTGCTCTT 1173
Qy 267 LeuArgGluHisPheIleLysAsnArgGlyIleThrProLysProSerLeuIleLysAla 286
Db 1174 ATGCTTAAC-----GCTAATCCAATCTCACCAATGATGCA-----AAAAAT 1218
Qy 287 AlaLeuIleAlaGlyAlaThrAspValGlyLeuGlyTyrProSerGlyAspGlnGlyTrp 306
Db 1219 ATAATAATGTCTACTGCAAAAAGCTGGGCGCCTCCAGCAAAACGTTGACTATGGTGCA 1278
Qy 307 GlyArgValThrLeuAspLysSerLeuAsnValAla----- 318
Db 1279 GGAAGATTGGATGCTGCTATGAGCTATAGGTTAGGTCAGGTAATTTTAGAGGAATAATATT 1338
Qy 319 -----TyrValAsnGluAlaThrAlaLeuThrThrGlyGlnLysAlaThr 333
Db 1339 GATGTACCAATCATTTATATTTTCAAGTTATCTGCCCGTCTCTCGTTACAGTGATACT 1398
Qy 334 TyrSerPheGln---ThrGlnAlaGlyLysProLeuLysIleSerLeuValTyrThrAsp 352
Db 1399 TGGACTTTTACGCAACAAATACAGTTATCCATCCGATCCGATTAATTAATTCCTGAC 1458
Qy 353 AlaProGlySerThrThrAlaSerTyrThrLeuValAsnAspLeuAspLeuValIleThr 372
Db 1459 -----TGGGCAAAATTACAAT-----CCCGATTTTGTATATATACCTCTAT 1497
Qy 373 AlaPro---AsnGlyGlnLysTyrValGly 381
Db 1498 GATCCGGCGCATGGAGGAAAGACCCCTGGG 1527

RESULT 14

ABX34289
 ID ABX34289 standard; DNA; 135638 BP.
 XX
 AC ABX34289;
 XX
 DT 11-FEB-2003 (first entry)
 XX
 DE S. atroolivaceus leinamycin biosynthesis gene cluster.
 XX
 KW Leinamycin biosynthesis gene cluster; Lnm; open reading frame; ORF;
 KW anti-tumour antibiotic; broad spectrum antimicrobial activity;
 KW Gram-positive; Gram-negative bacteria; chemical modification; metabolite;
 KW apo-carrier protein; holo-carrier protein; tumour; polyketide;
 KW hybrid polypeptide/polyketide metabolite; Lnm production; cytostatic;
 KW gene; ds.
 XX
 OS Streptomyces atroolivaceus.
 XX
 XX WO200277179-A2.
 XX
 XX 03-OCT-2002.
 XX
 XX 22-MAR-2002; 2002WO-US008937.
 XX
 XX 26-MAR-2001; 2001US-0278935P.
 XX
 XX (REGC) UNIV CALIFORNIA.
 XX (KYOW) KYOWA HAKKO KOGYO KK.
 XX
 XX Shen B, Cheng Y, Tang G;
 XX
 XX WPI; 2003-018907/01.
 DR P-PSDB; ABU11341, ABU11342, ABU11343, ABU11344, ABU11345, ABU11346,
 DR ABU11347, ABU11348, ABU11349, ABU11350, ABU11351, ABU11352, ABU11353,
 DR ABU11354, ABU11355, ABU11356, ABU11357, ABU11358, ABU11359, ABU11360,
 DR ABU11361, ABU11362, ABU11363, ABU11364, ABU11365, ABU11366, ABU11367,
 DR ABU11368, ABU11369, ABU11370, ABU11371, ABU11372, ABU11373, ABU11374,
 DR ABU11375, ABU11376, ABU11377, ABU11378, ABU11379, ABU11380, ABU11381,
 DR ABU11382, ABU11383, ABU11384, ABU11385, ABU11386, ABU11387, ABU11388,
 DR ABU11389, ABU11390, ABU11391, ABU11392, ABU11393, ABU11394, ABU11395,
 DR ABU11396, ABU11397, ABU11398, ABU11399, ABU11400, ABU11401, ABU11402,
 DR ABU11403, ABU11404, ABU11405, ABU11406, ABU11407, ABU11408, ABU11409,
 DR ABU11410, ABU11411.
 XX
 PT Novel gene cluster responsible for synthesis of leinamycin in
 PT Streptomyces atroolivaceus useful for making various peptide and/or
 PT polyketide, and/or hybrid polypeptide/polyketide metabolites.
 XX
 PS Claim 6; Page 81-127; 185pp; English.
 XX
 CC The present invention relates to the isolation of the Streptomyces
 CC atroolivaceus leinamycin (Lnm) biosynthesis gene cluster containing 71
 CC open reading frames (ORFs) (ORFs -35 through -1, ORFs LnmA through LnmZ,
 CC and ORFs +1 through +9). Leinamycin is a novel anti-tumour antibiotic
 CC produced by several Streptomyces species. It exhibits broad spectrum
 CC antimicrobial activity against Gram-positive and Gram-negative bacteria,
 CC but not against fungi. The polypeptides encoded by the Lnm biosynthesis
 CC gene cluster ORFs are useful for chemically modifying a molecule in a
 CC host cell. The host cell is a bacterium or eukaryotic cell, including a
 CC mammalian, yeast, plant, fungal, or insect cell. The molecule is an
 CC endogenous metabolite produced by the host cell or exogenously supplied
 CC metabolite, or an amino acid, and the polypeptide is a peptide synthetase
 CC or amino transferase. The polypeptides encoded by the Lnm gene cluster
 CC are useful for converting an apo-carrier protein to a holo-carrier
 CC protein. Lnm shows potent antitumour activity in tumour models in vivo.
 CC The Lnm gene cluster modules and/or catalytic domains are useful for
 CC making various peptide and/or polyketide, and/or hybrid
 CC polypeptide/polyketide metabolites. The proteins encoded by the ORFs are
 CC useful alone, or in combination with other active domains to modify
 CC various target substrates. The Lnm gene cluster is useful to upregulate
 CC endogenous Lnm production to permit Lnm production in cells and/or to

CC make various modified Lnm, Lnm, its analogue, or other polyketide,
 CC peptide or hybrid polyketide/peptide metabolites are useful as
 CC therapeutic agents, to treat a number of disorders, depending upon the
 CC type of metabolites. The present sequence represents the S. atroolivaceus
 CC leinamycin biosynthesis gene cluster
 XX
 SQ Sequence 135638 BP; 18570 A; 49096 C; 49039 G; 18933 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 8 34e-16 Length: 135638
 Score: 346.50 Matches: 140
 Percent Similarity: 42.39% Conservative: 55
 Best Local Similarity: 30.43% Mismatches: 154
 Indels: 111
 Query Match: 7
 Gaps: 23
 DB:
 US-09-985-689A-3 (1-433) x ABX34289 (1-135638)
 QY 6 GlyIleValIysAlaAspValAlaGlnAsnAsn----- 16
 DB 8067 GGCAGGGTGAAGCCGATCTGCGCCGACTCCACGCCAGATCGCGCGCAGAGGTATGG 8126
 QY 17 ---TyrGlyLeuTyrGlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThr 35
 DB 8127 GCGAGGGCGCACCGCGCAGGACGTCAAGTCTCGATGCTGCACAGCGCGCGGACACC 8186
 QY 36 GlyArgAsnAspSerSerMetHisGluAlaPheArgGlyLysIle-----ThrAlaLeu 53
 DB 8187 -----GAACACCCGAGACTCTGTCGGCAGGTCTCCGACGCGCGCAGC 8228
 QY 54 TyrAlaLeuGlyArgThrAsnAlaAsnAspProAsnGlyHisGlyThrHisValAla 73
 DB 8229 TTGTCCTCCCGCGAGGACGACATCGCC--GATTACACGCCCCACGCGCAGCTCGCC 8285
 QY 74 GlySerValLeuGly-----AsnAlaLeuAsnLysGlyMetAlaProGln 88
 DB 8286 TCGACCATCTCGCGCAGCGCGCGCTCCGACGCGCAAGAGCGGGTGTCTCGCTCGGC 8345
 QY 89 AlaAsnLeuValPheGlnSerIleMetAspSerSerGlyGlyLeuGlyLeuProSer 108
 DB 8346 GCGCGGTGTCTCGCGCAAGGTGCTCACTCCGAGGC-----AGCGCGCAGGATCG 8399
 QY 109 AsnLeuAsnThrLeuPheSerGlnAlaTTPAsnAla-----GlyAlaArgIleHis 125
 DB 8400 TGGATC-----ATCGCGGCATGGAGTGGCGCGCGCGCCAGCAGGCGCAGATCATC 8453
 QY 126 ThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrAlaAsn-----SerArg 142
 DB 8454 AGCATGAGCTGGGC-----GGCGCGGTGACCAAGAACACCGCATGACCGCAG 8501
 QY 143 GlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAlaGlyAsnGlu 162
 DB 8502 GCGTCGACCACTACGCCACGACACGCGCGCGCTTGTTCGTGATCGCGCGGCAACGCG 8561
 QY 163 GlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGly 182
 DB 8562 GCGCGCACTCC-----ATCAGACGCGCGCGCGCGCAGCTCCGCGTGCAGCGCGGC 8615
 QY 183 AlaThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsnProAsnHisIleAla 202
 DB 8616 GCC-----GTCGACTCCACCGCAGCTCGCC 8642
 QY 203 GlnPheSerSerArgGly-----AlaThrArgAspGlyArgIleLysProAspValThrAla 221
 DB 8643 GACTTCTCCAGCGAGGCGCGCGTGCAGCGCGCGCGCGGTGAAGCGCGGAGATCACCGCG 8702
 QY 222 ProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheThrAla 241
 DB 8703 CCGCGGCTCCACATCTGCGCGCGCGCGCTCCATCATCAGCGCGCGCTCCGCTAC----- 8756
 QY 242 AsnTyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleValAla 261
 DB 8757 -----TACACCAAGATGAGCGGCGCGCTCGATGCGCGCAGCGCGCGCTCGCC 8801

```
QY 262 GlyAsnValAlaGlnLeuArgLysPheIleLysAsnArgGlyIleThrProLysPro 281
DB 8802 GGTGTCCGCGGCTCTCCGCGGAGCACCACCGGAGTGCAGCGGACCGAGTCAAGGAG 8861
QY 282 SerLeuLysAlaAlaLeuLeuAlaGlyAlaThrAspValGlyLeuGlyTyrProSer 301
DB 8862 GCACGTGTCAGCAGCGCCCAAGCAACGCGCGG-----TACACCCG 8903
QY 302 GlyAspGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyrValAsn 321
DB 8904 TACCAGCGGCGCGCGCGCTCGACGCGCGCGCTGCACACACCGTCTTCGCC 8963
QY 322 GluAlaThrAlaLeuThrThrGlyGlnLysAlaThrTyrSerPheGlnThrGlnAlaGly 341
DB 8964 ACCACAGCGCCCTACTCC-----GGCTTCACACG----- 8993
QY 342 LysProLeuLysIleSerLeuValThrThrAspAlaProGlySerThr-----Thr 358
DB 8994 -----TGGCCCCCGAAGCCCGGAGACCGATGTCGGGACG 9029
QY 359 AlaSerTyrThrLeuValAsnAsp-----LeuAspLeuValIle-----Thr 372
DB 9030 GTACCGTACACCAACCTCGCGGACGCGCGCTCAACCTGCGCTCAACGCGCAC 9089
QY 373 AlaProAsnGlyGlnLysTyrValGlyAsnAspPheSerTyrProTyrAspAsnAsnTrp 392
DB 9090 GTCCCGCGCGGTG-----TTCAGGCTCTCCGAGGACCAT----- 9125
QY 393 AspGlyArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThrTyrThr 412
DB 9126 -----GTACCGTGCCTCCGCGCACCGGACCGGACCGGACCG 9155
QY 413 IleGluVal--GlnAlaTyrAsnValProSerGlyProGlnArgPheSerLeuAlaIle 431
DB 9156 GTCACCTGACCGCGGCTCTGGACAGCTGGCGGCGGACCGTGGTTCAGGCGCGGATC 9215

RESULT 15
AA161454
ID AA161454 standard; DNA; 2539 BP.
XX
AC AA161454;
DT 06-OCT-1997 (first entry)
DE
KW asymmetric hydrolase; dhpA; 4-substituted-1,4-dihydropyridine;
KW derivative; Streptomyces viridosporus; ester; chiral; synthesis;
KW cardiovascular; treatment; hypertension; ischaemic heart disease; ds.
OS Streptomyces viridosporus.
FH Key Location/Qualifiers
FT CDS 338..2539
FT /tag= a
FT /note= "no stop codon given"
FT misc_feature 950..2509
FT /tag= b
FT /note= "encodes AAW13666"
XX
PN WO9705243-A1.
XX
PD 13-FEB-1997.
XX
PF 30-JUL-1996; 96WO-JP002147.
XX
PR 31-JUL-1995; 95JP-00212975.
PR 29-FEB-1996; 96JP-00067478.
XX (SAOC) MERCIAN CORP.
XX
PI Arisawa A, Matsufuji M, Tsuruta T, Dobashi K, Nakashima T;
```

```
PI Isshiki K, Yoshioka T;
XX WPI; 1997-145682/13.
DR P-PSDB; AAW13666; AAW13667.
XX
PT Asymmetric hydrolase gene derived from Streptomyces viridosporus - acts
PT on 4-substituted-1,4-di:hydro:pyridine derivatives to produce chiral
PT derivatives useful for synthesis of cardiovascular drugs.
XX
XX Claim 3; Page 49-55; 78pp; Japanese.
XX
XX This sequence is the Streptomyces viridosporus dhpA gene which encodes an
CC asymmetric hydrolase which acts on 4-substituted-1,4-dihydropyridine
CC derivatives. The enzyme allows the efficient conversion of 4-substituted-
CC 1,4-dihydropyridine esters to chiral partially hydrolysed derivatives,
CC for use in the synthesis of cardiovascular drugs suitable for the
CC treatment of e.g. hypertension and ischaemic heart disease
XX
XX Sequence 2539 BP; 433 A; 959 C; 867 G; 280 T; 0 U; 0 Other;
XX
Alignment Scores:
Pred. No.: 1.15e-17 Length: 2539
Score: 343.50 Matches: 124
Percent Similarity: 38.76% Conservative: 45
Best Local Similarity: 28.44% Mismatches: 150
Query Match: 15.18% Indels: 117
DB: 2 Gaps: 17
US-09-985-689A-3 (1-433) x AAT61454 (1-2539)
QY 18 GlyLeuTyrGlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArg 37
DB 998 GGCTACGACGCAAGGCGGTGAAGATCGCCTCTGACACCGCGGTGTCGACAGCAGC--- 1054
QY 38 AsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGly 57
DB 1055 -----CATCCGACCTGAAGGCGCGGTGACCGGCTCCAAAGACTTCACC 1099
QY 58 ArgThrAsnAsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeu 77
DB 1100 GCGCGCGCGCGCGCGCGGACAGGTGGCGCCACGCGCACCGCTCGCTCGATCGCGCG 1159
QY 78 GlyAsnAlaLeuAsn-----LysGlyMetAlaProGlnAlaAsnLeuVal 92
DB 1160 GGCACGGCGCGCCAGTCCACAGGCAAGTACAAAGGCGCTGCACCGCGCGCGCATCTCTC 1219
QY 93 PheGlnSerIleMetAspSerSerGlyLeuGlyLeuGlyLeuProSerAsnAsnThr 112
DB 1220 AACGCCAAGGTCTCTGACGACTCC-----GGTTTCGGGAGCAGCTCCGGCATC 1267
QY 113 LeuPheSerGlnAlaTyrAsnAla-----GlyAlaArgIleHisThrAsnSerTyrGly 130
DB 1268 CTCGCGCGCATGGAGTGGCGCGCGCGCGCGCGCGCGCTCGTCAACATGAGCTGGGC 1327
QY 131 Ala-----ProValAsnGlyAlaTyrThrAlaAsnSerArgGln 143
DB 1328 GGCATGGACACACCGGAGACCGACCCGCTGGAGCGCGCG----- 1366
QY 144 ValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAlaGlyAsnGly 163
DB 1367 CTCGACAGCTCTCGCGGAGAGGCGCTCTGTTCGCGCATCGCGCGCGCAACGAGGCG 1426
QY 164 ProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAla 183
DB 1427 CCGGAGTCG-----ATCGGTTTCGCGCGGACCGCGCGCGCTCACGTCGCGCGC 1480
QY 184 ThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsnProAsnHisIleAlaGln 203
DB 1481 -----GTCGACGACAGGACAGCAAGTCGCCGAC 1507
QY 204 PheSerSerArgGly---AlaThrArgAspGlyArgIleLysProAspValThrAlaPro 222
DB 1508 TTCTCTCCACCGCGCGCGCGCTCGCGGACCGCGCGCATCAAGCGGCGGCTCACCGCTCCC 1567
```

| | | | |
|----|------|--|------|
| Qy | 223 | GlyThrPheIleuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsn | 242 |
| | | | |
| Db | 1568 | GGCGTGACATACAGCGCGCTCGCGAGGGCAACGACATCGCCAGGAGTGGGTGAG | 1627 |
| | | | |
| Qy | 243 | TyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGly | 262 |
| | | | |
| Db | 1628 | GGACCGCGCGCTACATGACCATCTCCGGCAGCTCGATGGCGACCCCGCACGTCGCGGGC | 1687 |
| | | | |
| Qy | 263 | AsnValAlaGlnLeuArgGluHisPheIleLysAsnArgLysIleThrProLysProSer | 282 |
| | | | |
| Db | 1688 | GGCGGGCCCTCTGAAGCAGCAG-----CACCCCGAC | 1720 |
| | | | |
| Qy | 283 | LeuIleLysAlaLeu-----IleAlaGlyAlaThrAspValGlyLeuGlyTyr | 299 |
| | | | |
| Db | 1721 | TGACCTCCGCCCAACTGAAGGGCGCGCTCACCGGCTCCACCAAGGCGGC-----AAGTAC | 1777 |
| | | | |
| Qy | 300 | ProSerGlyAspGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr | 319 |
| | | | |
| Db | 1778 | ACCCGCTCGAGCAGGGTTCCGGCCGATCCAGGCCGACAGGCGCTCCAG----- | 1828 |
| | | | |
| Qy | 320 | ValAsnGluAlaThrAlaLeuThrThrGlyGlnLysAlaThrTyrSerPheGlnThrGln | 339 |
| | | | |
| Db | 1829 | -----CAGACCGGTG | 1837 |
| | | | |
| Qy | 340 | AlaGlyLysProLeuLysIleSerLeu-----ValTrp-----ThrAspAla | 353 |
| | | | |
| Db | 1838 | ATCGCCACCGGTCCTCGGTGAGTTCGCGCTCCAGCAGTGGCGCACACCGACGACGAG | 1897 |
| | | | |
| Qy | 354 | ProGlySerThrThrAlaSerTyrThrLeuValAsnAspLeuAspLeuValIle----- | 371 |
| | | | |
| Db | 1898 | CCGTCACCAAGCAGCTGACTACCGCAACTCGGCACCCAGCAGCAGCTCAGCTGAAGCTG | 1957 |
| | | | |
| Qy | 372 | -----ThrAlaProAsnGlyGlnLysTyrValGlyAsnAspPheSerTyrPro | 387 |
| | | | |
| Db | 1958 | ACGTCGACCGCCACCGACCCCAAGGGCAAG----- | 1987 |
| | | | |
| Qy | 388 | TyrAspAsnAsnTyrAspGlyArgAsnAsnValGluAsnValPheIleAsnAlaProGln | 407 |
| | | | |
| Db | 1988 | -----GGGGCCCG | 1996 |
| | | | |
| Qy | 408 | SerGlyThrTyrThrIleGluValGlnAlaTyrAsnValProSerGly | 423 |
| | | | |
| Db | 1997 | GGGGCTTCTCAGCTGGGCGGCCACACAGCTGACCGTCCCGCGGC | 2044 |

Search completed: April 4, 2004, 01:11:31
Job time : 392.908 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 3, 2004, 23:50:44 ; Search time 63.4512 Seconds
(without alignments)
3787.066 Million cell updates/sec

Title: US-09-985-689A-3

Perfect score: 2263

Sequence: 1 NDVARGIVKADVAQNNYGLY.....EVQANNVPSGQRFSLAIVH 433

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5

Ygapop 10.0 , Ygapext 0.5

Fgapop 6.0 , Fgapext 7.0

Delpop 6.0 , Delext 7.0

Searched: 682709 segs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US09985689/runat_31032004_161807_4184/app.query.fasta_1.3498
-DB=issued_patents_NA -QFMT=fastap -SUFFIX=rn1 -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=BLOSUM62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09985689 @CGN 1 1.142 @runat_31032004_161807_4184 -NCPUS=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DSV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued_Patents_NA:**
1: /cgn2_6/prodata/2/ina/5A_COMB.seq:**
2: /cgn2_6/prodata/2/ina/5B_COMB.seq:**
3: /cgn2_6/prodata/2/ina/6A_COMB.seq:**
4: /cgn2_6/prodata/2/ina/6B_COMB.seq:**
5: /cgn2_6/prodata/2/ina/PCUS_COMB.seq:**
6: /cgn2_6/prodata/2/ina/backfiles1.seq:**

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|----|-------------------|
| 1 | 2040 | 90.1 | 3003 | 2 | US-08-873-479-41 |
| 2 | 1998.5 | 88.3 | 1920 | 4 | US-09-509-814A-3 |
| 3 | 1995.5 | 88.2 | 1923 | 4 | US-09-509-814A-7 |
| 4 | 1994.5 | 88.1 | 1923 | 4 | US-09-509-814A-5 |
| 5 | 443 | 19.6 | 1977 | 3 | US-08-894-818B-2 |
| 6 | 443 | 19.6 | 1977 | 4 | US-09-445-472-11 |
| 7 | 422.5 | 18.7 | 1236 | 3 | US-09-445-472-2 |
| 8 | 422.5 | 18.7 | 1566 | 4 | US-08-894-818B-4 |
| 9 | 422.5 | 18.7 | 1962 | 3 | US-08-894-818B-34 |
| 10 | 422.5 | 18.7 | 1962 | 4 | US-09-445-472-15 |
| 11 | 401.5 | 17.7 | 1977 | 3 | US-08-894-818B-6 |
| 12 | 348.5 | 15.4 | 2539 | 3 | US-09-000-016-3 |
| | | | | | Sequence 41, Appl |
| | | | | | Sequence 3, Appl |
| | | | | | Sequence 7, Appl |
| | | | | | Sequence 5, Appl |
| | | | | | Sequence 11, Appl |
| | | | | | Sequence 2, Appl |
| | | | | | Sequence 4, Appl |
| | | | | | Sequence 34, Appl |
| | | | | | Sequence 15, Appl |
| | | | | | Sequence 6, Appl |
| | | | | | Sequence 3, Appl |

Sequence 3, Appl
Sequence 1, Appl
Sequence 1, Appl
Sequence 4, Appl
Sequence 15, Appl
Sequence 2, Appl
Sequence 8, Appl
Sequence 5, Appl
Sequence 11, Appl
Sequence 14, Appl
Sequence 7, Appl
Sequence 11, Appl
Sequence 11, Appl
Sequence 11, Appl
Sequence 6, Appl
Sequence 11, Appl
Sequence 1, Appl
Patent No. 5336611
Patent No. 5217878
Sequence 1, Appl
Sequence 1, Appl
Sequence 1, Appl
Sequence 1, Appl
Sequence 1, Appl
Sequence 1, Appl
Sequence 72, Appl
Sequence 72, Appl
Sequence 72, Appl
Sequence 72, Appl
Sequence 12, Appl

US-09-514-340-3
US-09-000-016-1
US-09-514-340-1
US-07-671-3760-4
US-08-894-818B-15
US-08-750-532-2
US-08-750-532-8
US-08-894-818B-7
US-09-445-472-5
US-08-750-532-11
US-08-894-818B-14
US-08-750-532-7
US-08-322-677A-11
US-08-322-676-11
US-08-898-218-11
US-08-848-793-11
US-09-445-270-6
US-08-322-678-11
US-08-269-050-1
5336611-1
5217878-1
US-09-049-867-1
US-07-661-378A-1
US-07-918-318-1
US-08-413-724-1
US-08-853-494-1
US-08-875-154-1
US-07-706-691C-72
US-08-254-021-72
US-08-618-446-72
US-08-980-135-72
US-09-585-798-72
US-08-322-677A-12

ALIGNMENTS

RESULT 1

US-08-873-479-41
; Sequence 41, Application US/08873479
; Patent No. 5891701
; GENERAL INFORMATION:
; APPLICANT: Syma, Alan
; APPLICANT: Lynne, Christianson
; TITLE OF INVENTION: Nucleic Acids Encoding A Polypeptide
; TITLE OF INVENTION: Having Protease Activity
; NUMBER OF SEQUENCES: 57
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: No. 5891701c No. 5891701disk of No. 5891701th America
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/873,479
; FILING DATE: 12-JUN-1997
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Agilis, Cheryl H
; REGISTRATION NUMBER: 34,086
; REFERENCE/DOCKET NUMBER: 5251.000-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; TELEX:
; INFORMATION FOR SEQ ID NO: 41:
; SEQUENCE CHARACTERISTICS:

```
/ LENGTH: 3003 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: Genomic DNA
US-08-873-479-41

Alignment Scores:
Pred. No.: 5,06e-194 Length: 3003
Score: 2040.00 Matches: 385
Percent Similarity: 94.46% Conservative: 24
Best Local Similarity: 88.91% Mismatches: 24
Query Match: 90.15% Indels: 0
DB: 2 Gaps: 0

US-09-985-689A-3 (1-433) x US-08-873-479-41 (1-3003)

Qy 1 AsnAspValAlaArgGlyLeuValValAlaAspValAlaGlnAsnAsnTyrGlyLeuTyr 20
Db 1470 AATGACGTGGCGGTGGCATTGTGAAGCAGACGTCGCACAAATAAATCTTTGGCTTATAT 1529

Qy 21 GlyGlnGlyGlnValValAlaAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 1530 GCACAGGACAGATGTAGCATGTCTGATCTCTGGCTTGATACAGGAAGAAATGACAGT 1589

Qy 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyValThrAsn 60
Db 1590 TCGATCGATGACGATTCGCGGTGAAGATTACGACATATATGCTGGCAGAGCAAT 1649

Qy 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnAla 80
Db 1650 AACGCCAATGATCCAAATGGACATGGAAACCAATGTCTGGTATCTGTGTAGGAATGCT 1709

Qy 81 LeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSerSer 100
Db 1710 ACAATTAAGGATGACGCGGACCAATCTAGTCTTTCATCTATATATGATAGTGGT 1769

Qy 101 GlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaThrAsnAla 120
Db 1770 GGAGGGCTGGAGGACTACCTGTAATCTACAAACATTTATTCAGTCAAGCATATAGTGT 1829

Qy 121 GlyAlaArgIleHisThrAsnSerTyrGlyAlaProValAsnGlyValAlaTyrThrAlaAsn 140
Db 1830 GGAGCGAGATTCATAGCAATTCATGGGGGCTCCAGTAAACGGTGCCTATACGACAGAC 1889

Qy 141 SerArgGlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAlaGly 160
Db 1890 TCTCGAATGCTGATGATATGTGAGAAAAATGATATGACGATCTTTTTCGGCGCGGA 1949

Qy 161 AsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThr 180
Db 1950 AATGAGGACCGAGTACGATGATCAATCAGTGCACCGAGACAGCAAAAATGCGATTACA 2009

Qy 181 ValGlyAlaThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsnProAsnHis 200
Db 2010 GTTGGGCAACCGAAACCTACGTCCAAAGCTTCGGATCTTATCGGATAATATTAACCAT 2069

Qy 201 IleAlaGlnPheSerArgGlyAlaThrArgAspGlyArgIleLysProAspValThr 220
Db 2070 GTTCTCAATTCCTTCCAGGAGTCTTACTAGAGATGGAGTATTAAAGCGGACGTCATG 2129

Qy 221 AlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTyr 240
Db 2130 GCACCGGTACGTATATCTCTCTGTCTAGATCATCATAGTCTCCAGATTCCTCATCTGG 2189

Qy 241 AlaAsnTyrAsnSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIleVal 260
Db 2190 GCAACCATGATAGTAATATGCTACATGGGTGATCTTCTATGGCTACTCCAAATGTA 2249

Qy 261 AlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrProLys 280
Db 2250 GCAGGTATGTGGACATTAAGGAGCATTTTGTGAAAATAGAGGGGTAACTCTCTAAG 2309

Qy 281 ProSerLeuIleLysAlaAlaLeuIleAlaGlyAlaThrAspValGlyLeuGlyTyrPro 300
Db 2310 CTTCCCTTTTAAAGCTGCTTTAAATGCGAGTGTGCGGATGTTGGACTTGCCTTCCA 2369

Qy 301 SerGlyAspGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyrVal 320
Db 2370 AATGGTAACCAAGGATGGGAAGAGTAACGTTAGATAAATCCCTAAATGTCGATTTGTG 2429

Qy 321 AsnGluAlaThrAlaLeuThrThrGlyGlnLysAlaThrTyrSerPheGlnThrGlnAla 340
Db 2430 AATGAAACGAGCCCTTTATCAACAGTCAAAAAGCAACATATTCGTTTACGGCTCAAGCT 2489

Qy 341 GlyLysProLeuLysIleSerLeuValTyrThrAspAlaProGlySerThrThrAlaSer 360
Db 2490 GGTAAACCTTAAATAATATCATTGTTGGTCAGATGCACAGGTAGCAGCGGCATCA 2549

Qy 361 TyrThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyGlnLysTyrVal 380
Db 2550 CTAACTTTAGTGAATGATTAGACTTTAGTAATCACTGCACCAATGGAATGAACTAAATACGTC 2609

Qy 381 GlyAsnAspPheSerTyrProTyrAspAsnAsnTyrAspGlyArgAsnAsnValGluAsn 400
Db 2610 GGAATGACCTTTACAGCAGCGTATGATAACAATTTGGATGGCAGAAACAACTGGGAAAT 2669

Qy 401 ValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsnVal 420
Db 2670 GTGTTTATCAATGCTCTCTCAAGCGGAACGTTATACAGTCAAGTGCAGGCTTACAATGTA 2729

Qy 421 ProSerGlyProGlnArgPheSerLeuAlaIleValHis 433
Db 2730 CCAGTAAGTCCGCAACCTTTTCTTTAGCGATGTACAT 2768

RESULT 2
US-09-509-814A-3
; Sequence 3, Application US/09509814A
; Patent No. 6376227
; GENERAL INFORMATION:
; APPLICANT: TAKAWA, MIKIO
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SAEKI, KATSUHIKA
; APPLICANT: KUBOTA, HIROMI
; APPLICANT: HITOMI, JUN
; APPLICANT: KAGEYAMA, YASUSHI
; APPLICANT: SHIKATA, SHITSUW
; APPLICANT: NOMURA, WASAFUMI
; TITLE OF INVENTION: ALKALINE PROTEASE
; FILE REFERENCE: 0327-0832-0PCT
; CURRENT APPLICATION NUMBER: US/09/509,814A
; CURRENT FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: PCT/JP98/04528
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: JP 9-274570
; PRIOR FILING DATE: 1997-06-08
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 3
; LENGTH: 1920
; TYPE: DNA
; ORGANISM: Bacillus sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1920)
US-09-509-814A-3

Alignment Scores:
Pred. No.: 3,61e-190 Length: 1920
Score: 1998.50 Matches: 382
Percent Similarity: 94.01% Conservative: 26
Best Local Similarity: 88.02% Mismatches: 25
Query Match: 88.31% Indels: 1
DB: 4 Gaps: 1

US-09-985-689A-3 (1-433) x US-09-509-814A-3 (1-1920)
```


| | | | |
|----|------|---|------|
| Qy | 80 | AlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer | 99 |
| Db | : | : | : |
| Db | 859 | TCCACTAATAAGGAATGGCGCCTCAGCGGAATCTAGTCTTCCAATCTATCATGGATAGC | 918 |
| Qy | 100 | SerGlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTyrAsn | 119 |
| Db | : | : | : |
| Db | 919 | GGTGGGGGACTTGGAGGACTACCTTCCGAACTCTGCAAACTTATTCAGCCAAGCATACAGT | 978 |
| Qy | 120 | AlaGlyAlaArgIleHisThrAsnSerTyrGlyAlaProValAsnGlyAlaTyrThrAla | 139 |
| Db | : | : | : |
| Db | 979 | GCTGGTCCCAAGAAATTCACAAATCTCTGGGAGCAGCAGTGAATGGGGCTTACACAACA | 1038 |
| Qy | 140 | AsnSerArgGlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAla | 159 |
| Db | : | : | : |
| Db | 1039 | GATTCCAGAATGTGGATGACTATGTGGCAAAATGATATGACGATCCCTTTTCGCTGCC | 1098 |
| Qy | 160 | GlyAsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle | 179 |
| Db | : | : | : |
| Db | 1099 | GGGAATGAAGACCCGAACGGCGGAACCATCAGTGCACCGACAGCTAAAAATGGCAATA | 1158 |
| Qy | 180 | ThrValGlyAlaThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsnProAsn | 199 |
| Db | : | : | : |
| Db | 1159 | ACAGTCGAGCTACGGAAACCTCGGCCCAAGCTTTGGGCTTATGCGGCAATATCCAC | 1218 |
| Qy | 200 | HisIleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspVal | 219 |
| Db | : | : | : |
| Db | 1219 | CATGTGCACAGATTCTCTTACGTGGACCGACGAAGGATGACCGATCAAAACGGATGTC | 1278 |
| Qy | 220 | ThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe | 239 |
| Db | : | : | : |
| Db | 1279 | ATGGCACCGGAAGCTTCATCTATCAGCAAGATCTTCTCTTGGCACCGGATTCCTCCTTC | 1338 |
| Qy | 240 | TrpAlaAsnTyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle | 259 |
| Db | : | : | : |
| Db | 1339 | TGGCGAACCATGACAGTAATATGCATACATGGGTGGAACGTCCATGGGTACACCGATC | 1398 |
| Qy | 260 | ValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro | 279 |
| Db | : | : | : |
| Db | 1399 | GTTGCTGAAACGTGGCACAGCTTCGTGAGCATTTGTGAAAAACAGAGGCATCACACCA | 1458 |
| Qy | 280 | LysProSerLeuIleLysAlaAlaLeuIleAlaGlyAlaThrAspValGlyLeuGlyTyr | 299 |
| Db | : | : | : |
| Db | 1459 | AAGCCTTCTCTATTAAAGCGCACTGATTCGCGTGCAGCTGACATCGGCTTGGGTAC | 1518 |
| Qy | 300 | ProSerGlyAspGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr | 319 |
| Db | : | : | : |
| Db | 1519 | CCGAACGGTAACCAAGGATGGGACGAGTGATCTGGATAAATCCCTGAACGTTGGCTAT | 1578 |
| Qy | 320 | ValAsnGluAlaThrAlaLeuThrThrGlyGlnLysAlaThrTyrSerPheGlnThrGln | 339 |
| Db | : | : | : |
| Db | 1579 | GTGAACGAGTCCAGTCTCTATCCACCACGCCAAAGCGACGTACTCGTTTACTGCTACT | 1638 |
| Qy | 340 | AlaGlyLysProLeuLysIleSerLeuValThrThrAspAlaProGlySerThrThrAla | 359 |
| Db | : | : | : |
| Db | 1639 | GCCGCGAAGCCTTTGAAATCTCCCTGGTATGGTCTGATGCCCTGGGAGCACCACTGCT | 1698 |
| Qy | 360 | SerTyrThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyGlnLysTyr | 379 |
| Db | : | : | : |
| Db | 1699 | TCCGTAACGCTGTGCAATGATCTGGACCTTGTCTATTACCGCTCCCAATGGCACAGTAT | 1758 |
| Qy | 380 | ValGlyAsnAspPheSerTyrProTyrAspAsnAsnThrAspGlyArgAsnAsnValGlu | 399 |
| Db | : | : | : |
| Db | 1759 | GTAGGAAATGACTTTTACTTCGCCATACAAATGATTAACCTGGGATGGCCCAATTAACGTGAA | 1818 |
| Qy | 400 | AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn | 419 |
| Db | : | : | : |
| Db | 1819 | AATGTATTATTATATGCACCACCAAGCGGACGTATACAAATTGAATGACAGCTTTAAC | 1878 |
| Qy | 420 | ValProSerGlyProGlnArgPheSerLeuAlaIleValHis | 433 |
| Db | : | : | : |
| Db | 1879 | GTACCGGTTGGACCACAGAAGCTTCTCGTTGGCAATTGTGAAT | 1920 |

RESULT 4

```

US-09-509-814A-5
; Sequence 5, Application US/095098.4A
; Patent No. 6376227
; GENERAL INFORMATION:
; APPLICANT: TAKAIWA, MIKIO
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SAEKI, KATSUYOSHI
; APPLICANT: KUBOTA, HIROMI
; APPLICANT: HITOMI, JUN
; APPLICANT: KAGEYAMA, YASUSHI
; APPLICANT: SHIKATA, SHITSUW
; APPLICANT: NOMURA, MASAFUMI
; TITLE OF INVENTION: ALKALINE PROTEASE
; FILE REFERENCE: 0327-0832-0PCT
; CURRENT APPLICATION NUMBER: US/09/509,814A
; CURRENT FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: PCT/JP98/04528
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: JP 9-274570
; PRIOR FILING DATE: 1997-06-08
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 5
; LENGTH: 1923
; TYPE: DNA
; ORGANISM: Bacillus sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1923)
US-09-509-814A-5

Alignment Scores:
Score: 9.1e-190
Percent Similarity: 1994.50
Best Local Similarity: 94.01%
Query Match: 87.56%
Indels: 88.14%
Gaps: 4

```

QY 160 GlyAsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 179
Db 1099 GGGATGAGGACCGAAGCGGCGAACCATCATGTCACCGACACAGTAAATGCAATA 1158
QY 180 ThrValGlyAlaThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsnProAsn 199
Db 1159 ACAGTCGAGACTACGGAACCTCGGCCAAGCTTGGTCTTATGCGCAATATCAAC 1218
QY 200 HisIleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspVal 219
Db 1219 CATGTGGCAGAGTTCTCTTCAGTGGACCGCAAGAGATGGACGATCAACCGGATGTC 1278
QY 220 ThrAlaProGlyThrPheIleLeuSerAlaArgSerLeuAlaProAspSerSerPhe 239
Db 1279 ATGGACCGGACGTTCTACTATCAGCAAGATCTTCTTGACCGGATCTCTCTTC 1338
QY 240 TrpAlaAsnTyrAsnSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIle 259
Db 1339 TGGCGGAACCATCAGCAGTAATATGATACATGATGGTGGACGTCATGGCTACACCGATC 1398
QY 260 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro 279
Db 1399 GTTGTGGAACGTGGCAGCTTCTGAGCATTTTGTGAACAGAGGATCACACCA 1458
QY 280 LysProSerLeuIleLysAlaAlaLeuIleAlaGlyAlaThrAspValGlyLeuGlyTyr 299
Db 1459 AAGCTTCTCTATTAAAGCGCACTGATTCGCGTGCAGCTGACATCGGCTTGGCTAC 1518
QY 300 ProSerGlyAspGlnGlyTrpGlyArgValThrIleAspLysSerLeuAsnValAlaTyr 319
Db 1519 CCGAAGCGTAACCAAGATGGGAGCGAGTGGATTTGGATAATCCTCGAAGCTTGGCTAT 1578
QY 320 ValAsnGluAlaThrAlaLeuThrThrGlyGlnLysAlaThrTyrSerPheGlnThrGln 339
Db 1579 GTGACGAGTCCAGTTCTCTATCCACCGCAAGCAAGGACGATCTCGTTACTGCTACT 1638
QY 340 AlaGlyLysProLeuLysIleSerLeuValThrThrAspAlaProGlySerThrAla 359
Db 1639 GCGGCAAGCCTTTGAAATCTCCCTGGTATGCTGATGCCCGCGAGCACAACTGCT 1698
QY 360 SerTyrThrLeuValAsnAspLeuValIleThrAlaProAsnGlyGlnLysTyr 379
Db 1699 TCGTAACGCTTGTCATGATCTGGACCTTGTCATACCGTCCCAATGGCACACAGTAT 1758
QY 380 ValGlyAsnAspPheSerTyrProTyrAspAsnAsnTyrAspGlyArgAsnValGlu 399
Db 1759 GTAGGAATGACTTTTCTTCGCATACCAATGATAACTGGGATGGCGCAATAACGTAGAA 1818
QY 400 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 419
Db 1819 AATGATTTATTATGACCAACAAAGCGGAGCTATACANTGAGGTACAGGCTTATTAAC 1878
QY 420 ValProSerGlyProGlnArgPheSerLeuAlaIleValHis 433
Db 1879 GTACCGTTGGACACAGACCTTCTCTGTTGCAATTTGTAAT 1920

STREET: 419 Seventh Street N.W., Ste. 300
CITY: Washington
STATE: D.C.
COUNTRY: United States of America
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/894,818B
FILING DATE: 20-MAY-1998
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP96/03253
FILING DATE: 07-NOV-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 323285/1995
FILING DATE: 12-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Browdy, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: TAKAKURA=1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-5197
TELEFAX: (202) 737-3528
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1977 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
US-08-894-818B-2

Alignment Scores:
Pred. No.: 1,43E-34 Length: 1977
Score: 443.00 Matches: 137
Percent Similarity: 46.03% Conservative: 66
Best Local Similarity: 31.07% Mismatches: 170
Query Match: 19,58% Indels: 68
DB: 3 Gaps: 16

US-09-985-689A-3 (1-433) x US-08-894-818B-2 (1-1977)

QY 8 ValLysAlaAspValAlaGlnAsnAsnTyrGlyLeuTyrGlyGlnGlyValValAla 27
Db 433 ATAGGGCGCGATACCGTCTGGAACCTCCTCGGCTACGACGAAGCGTGTGGTGGTGC 492
QY 28 ValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGluAlaPheArg 47
Db 493 ATGTCGATACGGGTATAGCGGAAC-----CACCCGATCTGAAG 534
QY 48 GlyLysIleThrAlaLeuTyr---AlaLeuGlyArgThrAsnAsnAlaAsnAspProAsn 66
Db 535 GGCAAGGTCATAGGCTGTGACGACGCGCTCAACGGCAGGTCCGACCCCTACGATGACCAG 594
QY 67 GlyHisGlyThrHisValAlaGlySerValLeuGly-----AsnAlaLeuAsn 82
Db 595 GGAACGGAACCCACCGTTGGGTATCTGTCGCGAACCGGACGCGTAACTCCAGTAC 654
QY 83 LysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerSer 100
Db 655 ATAGGCGTCCCGCGCGGAAGCTCGTGGCGTCAAGTCTCGTGGCGGACGCGTTCG 714
QY 101 GlyGlyLeuGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsnAla 120
Db 715 GGAAGCGTCTCCACCATCATCGCGGGGTGTGACTGGTCTCGTCAGAACCAAGGACAGTAC 774
QY 121 GlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrAlaAsn 140
Db 775 GGGATAGGTCATCAACCTCTCCCTCGGCTCTCTCCAGAGCTCCCGACGGAACCGACTCC 834

APPLICANT: ASADA, Kiyozo
 APPLICANT: KATO, Ikunoshin
 TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
 FILE REFERENCE: TAKAKURA=6
 CURRENT APPLICATION NUMBER: US/09/445,472
 CURRENT FILING DATE: 1999-12-06
 PRIOR APPLICATION NUMBER: 151989/1997
 PRIOR FILING DATE: 1997-06-10
 NUMBER OF SEQ ID NOS: 33
 SOFTWARE: Patent in version 3.0
 SEQ ID NO 11
 LENGTH: 1977
 TYPE: DNA
 ORGANISM: Artificial Sequence
 FEATURE:
 OTHER INFORMATION: Synthetic
 US-09-445-472-11
 Alignment Scores:
 Pred. No.: 1,43e-34 Length: 1977
 Score: 443.00 Matches: 137
 Percent Similarity: 46.03% Conservative: 66
 Best Local Similarity: 31.07% Mismatches: 170
 Query Match: 19.58% Indels: 68
 DB: 4 Gaps: 16
 US-09-985-689a-3 (1-433) x US-09-445-472-11 (1-1977)
 QY 8 VallysAlaAspValAlaGlnAsnAsnTyGlyLeuTyGlyGlnValValAla 27
 Db 433 ATAGGGCGGATACCGCTCGGAACCTCCCTCGCTACCGAAGCGGTGGTGGTGGC 492
 QY 28 ValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGluAlaPheArg 47
 Db 493 ATCTCGATACGGGTATAGACCGAAC-----CACCCGATCTGAAG 534
 QY 48 GlyIleThrAlaLeuTy-----AlaLeuGlyArgThrAsnAsnAlaAsnAspProAsn 66
 Db 535 GGCAGGTCTATAGCTGTCGACGCGCTCAACGGCAGGTGACCCCTACGATGACGAG 594
 QY 67 GlyHisGlyThrHisValAlaGlySerValLeuGly-----AsnAlaLeuAsn 82
 Db 595 GGACACGAGAACCCAGCTTGGGTATGCTTCCCGGAACCGCAGCGGTAACTCCAGTAC 654
 QY 83 LysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerSer 100
 Db 655 ATAGCGTCTGCGCGCGCGGAGCTCTCGCGCTCAAGGTCTCTCGGTGCGCAGCGGTTCG 714
 QY 101 GlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsnAla 120
 Db 715 GGAAGCGTCTCCACCATATCGCGGTGTGACTGGTCTGCTCCAGAACAGCAAGTAC 774
 QY 121 GlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyThrAlaAsn 140
 Db 775 GGGATAAGGGTCTCAACCTCTCCCTCGCTCTCCAGAGCTCCGACGCGAACCGACTCC 834
 QY 141 SerArgGlnValAspGluTyValArgAsnAsnAspMetThrValLeuPheAlaAlaGly 160
 Db 835 CTCAGTCAGCGCGTCAACACCGCTGGGACCGCGGTATAGTAGTCTGCTGCGCGCGCGC 894
 QY 161 AsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThr 180
 Db 895 AACAGCGGCGCGAACACCTACCTCGCTCACCCTCGCTCACCCTCGCTCACCCTCGCTC 954
 QY 181 ValGlyAlaThrGluAsnTyArgProSerPheGlySerIleAlaAspAsnProAsnHis 200
 Db 955 GTCGGTGCA-----GTTGACAGCAACGACAC 981
 QY 201 IleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspValThr 220
 Db 982 ATCCGACGCTTCTCCAGCAGGAGGACCGACCGCGGAGCGAAGGCTCAAGCGGAGTCTGC 1041
 QY 221 AlaProGlyThrPheIleLeuSerAlaArgSerLeuAlaProAspSerSerPheTrp 240

141 SerArgGlnValAspGluTyValArgAsnAsnAspMetThrValLeuPheAlaAlaGly 160
 Db 835 CTCAGTCAGCGCGTCAACACCGCTGGGACCGCGGTATAGTAGTCTGCTGCGCGCGCGC 894
 QY 161 AsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThr 180
 Db 895 AACAGCGGCGCGAACACCTACCTCGCTCACCCTCGCTCACCCTCGCTCACCCTCGCTC 954
 QY 181 ValGlyAlaThrGluAsnTyArgProSerPheGlySerIleAlaAspAsnProAsnHis 200
 Db 955 GTCGGTGCA-----GTTGACAGCAACGACAC 981
 QY 201 IleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspValThr 220
 Db 982 ATCCGACGCTTCTCCAGCAGGAGGACCGACCGCGGAGCGAAGGCTCAAGCGGAGTCTGC 1041
 QY 221 AlaProGlyThrPheIleLeuSerAlaArgSerLeuAlaProAspSerSerPheTrp 240
 Db 1042 GCCCGCGCGCTTACATCATAGCCCGCGCGCGCAGC-----GGAACGATGAGCGC 1092
 QY 241 AlaAsnTyArgSerIleAlaTyMetGlyGlyThrSerMetAlaThrProIleVal 260
 Db 1093 ACCCGGATACGACTACTACCAAGGCTCTGGAACGAGCATGCGCCACCGCGACGTT 1152
 QY 261 AlaGlyAsnValAlaGlnLeuArgGluHisPheIleIleIleIleIleIleIleIleIle 279
 Db 1153 TCGGCGCTTCCGCGCTCATCTCCAG-----GCCACCGAGCTGGACCGCGAC 1203
 QY 280 -----LysProSerLeuIleLysAlaAlaLeuIleAlaGlyAlaThrAspValGly 296
 Db 1204 AAGGTGAAGACCGCTCATCGAGCGCGCATATAGTCCGCCCAAGGAGATAGCGGAC 1263
 QY 297 LeuGlyTyProSerGlyAspGlnGlyTyArgValThrLeuAspLysSerLeuAsn 316
 Db 1264 ATCGCTTAC-----GGTGGGTAGGTGAACGTCTACAGGCC-----Lys 331
 QY 317 ValAlaTyValAsnGluAlaThrAlaLeuThrThrGlyGln-----Lys 331
 Db 1303 ATCAAGTACGACGACTACGCAAGCTCACCTTCCACCGCTCGTCCGCGCAAGGAGAC 1362
 QY 332 AlaThrTySerPheGlnThrGlnAlaGlyLysProLeuIleIleSerLeuValTrpThr 351
 Db 1363 GCCACCCACACTTCGAGCTCAGCGGCGCGCCACTTCTGTCGCGCCACCTCTACTGGGAC 1422
 QY 352 AspAlaProGlySerThrThrAlaSerTyThrLeuValAsnAspLeuValIle 371
 Db 1423 -----ACGGCTCGAGCGACATCGACCTTACCTCTACCTC 1452
 QY 372 ThrAlaProAsnGlyGlnLysTyValGlyAsnAspPheSerTyProTyArgAspAsn 391
 Db 1453 TACGACCCCAAC-----GGGAACGAGGTGACTACTCTCTACACCGCCTAC 1497
 QY 392 TrpAspGlyArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThrTy 411
 Db 1498 TAC-----GGTTCGAGAAGTCTGCTACTACACCGCGCGGAGTCTGCTGCTGCTG 1545
 QY 412 ThrIleGluValGlnAlaTyAsnValProSerGlyProGlnArgPheSerLeuAlaIle 431
 Db 1546 ACGGTCAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1596
 QY 432 Val 432
 Db 1597 GTC 1599
 RESULT 6
 US-09-445-472-11
 ; Sequence 11, Application US/09445472
 ; Patent No. 6358728
 ; GENERAL INFORMATION:
 ; APPLICANT: TAKAKURA, Hikaru
 ; APPLICANT: MORISHITA, Mio
 ; APPLICANT: SHIMOJO, Tomoko

Db 1042 GCGCCGGGCTTGACATCATAGCCCGCGGCAGC-----GGACAGCATGGGC 1092
Qy 241 AlaAsnTyrAsnSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProLeuVal 260
Db 1093 ACCCGATAAAGCACTACTACACCAAGGCTCTGGACACCATGGCCACCCGACGTT 1152
Qy 261 AlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyLeuPro--- 279
Db 1153 TCGGGCGTGGCGGCTCATCTCCAG-----GCCACCCGAGTGGACCCCGGAC 1203
Qy 280 -----LysProSerLeuIleLysAlaLeuIleAlaGlyAlaThrAspValGly--- 296
Db 1204 AAGGTGAAGACCGCCCTCATCGAGACCGCCGACATAGTCGCCGCCCAAGAGATAGCGGAC 1263
Qy 297 LeuGlyTyrProSerGlyAspGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsn 316
Db 1264 ATCGCCTAC-----GGTGGGTAGGTGAAGCTCTACAGGCC----- 1302
Qy 317 ValAlaTyrValAsnGluAlaThrAlaLeuThrThrGlyGln-----Lys 331
Db 1303 ATCAAGTACGACGACTACGCCAAGCTCACCTTCACCGGCTCGCGGACCAAGGAAGC 1362
Qy 332 AlaThrTyrSerPheGlnThrGlnAlaGlyLysProLeuLysIleSerLeuValThrThr 351
Db 1363 GCCACCCACACCTTCGACGTACGGCGGCCACCTTCGTGACCCGCCACCTCTACTGGGAC 1422
Qy 352 AspAlaProGlySerThrThrAlaSerTyrThrLeuValAsnAspLeuValIle 371
Db 1423 -----ACGGGCTCGAGCGACATCGACCTCTACCTC 1452
Qy 372 ThrAlaProAsnGlyGlnLysTyrValGlyAsnAspPheSerTyrProTyrAspAsnAsn 391
Db 1453 TAGGACCCCAAC-----GGGAACGAGGTGACTACTCTCTACACCCGCCCTAC 1497
Qy 392 TrpAspGlyArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThrTyr 411
Db 1498 TAC-----CGCTTCGAGAGGTGCGGTACTACACCCCGCGCGGACCTGG 1545
Qy 412 ThrIleGluValGlnAlaTyrAsnValProSerGlyProGlnArgPheSerLeuAlaIle 431
Db 1546 ACGGTCAAGGTCTGCTAGCTAC-----AAGGGCGCGCGGAACCTTACCAAGTCAAGCTC 1596
Qy 432 Val 432
Db 1597 GTC 1599

RESULT 7
US-09-445-472-2
; Sequence 2, Application US/09445472
; Patent No. 6358726
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA-6
; CURRENT APPLICATION NUMBER: US/09/445,472
; CURRENT FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 2
; LENGTH: 1236
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-09-445-472-2

Alignment Scores: 7.84e-33 Length: 1236
Pred. No.: 422.50 Matches: 138
Score: 43.96% Conservativity: 55
Best Local Similarity: 31.44% Mismatches: 167
Query Match: 18.67% Indels: 79
DB: 4 Gaps: 17
US-09-985-689A-3 (1-433) x US-09-445-472-2 (1-1236)
Qy 16 AsnTyrGlyLeuTyrGlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThr 35
Db 58 AACTGGGATATGATGCTTCTGAATACACATAGGAATTAATGACACTGGAATGAC--- 114
Qy 36 GlyArgAsnAspSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAla 55
Db 115 -----GCTTCTCATCCAGATCTCCAAGGAAAGTA----- 144
Qy 56 LeuGlyArgThrAsnAsnAlaAsn-----AspProAsnGlyHisGly 69
Db 145 ATTGGGGTGGTATGATTTTTCATGTGTAGGATTTATCCATGATGACCATGACATGGA 204
Qy 70 ThrHisValAlaGlySerValLeuGly-----AsnAlaLeuAsnLysGly 84
Db 205 ACTCATGTAGCTTCAATAGCAGCTGCTACTGGAGCAGCAAGTAATGCAAGTACAAGGA 264
Qy 85 MetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerSerGlyGly 102
Db 265 ATGCTCTCAGGAGCTAGCTGGCGGGAATTAAGGTTCTAGGTCCGATGTTCTTGGAGC 324
Qy 103 LeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsnAlaGlyAla 122
Db 325 ATATCTACTAATTAAGGAGTGTAGTGGCGCTTGCATTAACAAGATAAGTACGGAATT 384
Qy 123 ArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrAlaAsnSerArg 142
Db 385 AAGGTCAATTAATCTTCTCTTGGTTCAAGCCAGAGCTCAGATGCTACTGCGTCTTAAGT 444
Qy 143 GlnValAspGlnTyrValArgAsnAsnAspMetThrValLeuPheAlaAlaGlyAsnGlu 162
Db 445 CAGCTCTTAATGACGCGTGGGATGCTGGATTAAGTGTGTTGCTGCGCTGGAACAGT 504
Qy 163 GlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGly 182
Db 505 GGACCTAACAAAGTATACAAATCGGTTCTCCAGCAGCTGCAAGCAAAAGTTATTACAGTGA 564
Qy 183 AlaThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsnProAsnHisIleAla 202
Db 565 GCCCTTGACCAAGTAT-----GATGTTATACA 591
Qy 203 GlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspValThrAlaPro 222
Db 592 AGCTTCTCAAGCAGAGGGCCAACTGTCAGACGCGGAGCTTAAGCTGAGGTGTTGCTCCA 651
Qy 223 GlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsn 242
Db 652 GGAACCTGGATTAATGCTGCCAGAGCAAGT-----GGAACCTAGCATGGGTCAACCA 702
Qy 243 TyrAsnSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIleValAlaGly 262
Db 703 ATTAATGACTATTACACAGCAGCTCTGGGACATCAATGCAACTCTCCTCAGTAGCTGT 762
Qy 263 AsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro----- 279
Db 763 ATTGACGCCCTCTTGTCTCCAA-----GCACACCCGAGCTGGACTCCAGACAAAGTA 813
Qy 280 LysProSerLeuIleLysAlaAlaLeuIleAlaGlyAlaThrAspValGly---LeuGly 298
Db 814 AAAACAGCCCTCATAGAACTGCTGATATCGTAAGCCAGATGAATAGCCGATATAGCC 873
Qy 299 TyrProSerGlyAspGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAla 318
Db 874 TAC-----GGTGCAGGTAGGTTAATGCATACAGGCTATAAAC----- 912

QY 319 TyrValAsnGluAlaThrAlaLeuThrThrGlyGlnLysAla-----Thr 333
Db 913 TAGGATAACTATCAAGTAGTGTCTACTGGATATGTTGCCAACAAAGGACGCCAACT 972
QY 334 TyrSerPheGlnThrGlnAlaGlyLysProLeuLysIleSerLeuValThrThrAspAla 353
Db 973 CACCAGTTCGTTATTAGCGAGCTTCGTTCACTGCACTGACCAATTATCTGGGACCAATGCC 1032
QY 354 ProGlySerThrThrAlaSerThrLeuValAsnAspLeuAspLeuValIleThrAla 373
Db 1033 AAT-----AGCGACTTGATCTTACTCTACGAT 1062
QY 374 ProAsnGlyGlnLysTyrValGlyAsnAspPheSerTyrProTyrAspAsnThrPasp 393
Db 1063 CCCAATGGAAACCAAGTT-----GACTACTCTTAC-----ACC 1095
QY 394 GlyArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThrThrIle 413
Db 1096 GCTACTATGATTCGAAAGGTTGTTATTACAAACCACTGATGGAACATGGCAATT 1155
QY 414 GluValGlnAlaTyrAsnValProSerGlyProGlnArgPheSerLeuAlaIleVal 432
Db 1156 AAGGTTGTAAGCTAC-----AGCGGAAGTGCAAACTATCAAGTAGATGTGTA 1203

RESULT 8

US-08-894-818B-4
; Sequence 4, Application US/08994818B
; Patent No. 6261922
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: YAMAMOTO, Katsuhiko
; APPLICANT: MITTA, Masanori
; APPLICANT: ASADA, Kiyozo
; APPLICANT: TSUNASAWA, Susumu
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Browdy and Neimark
; STREET: 419 Seventh Street N.W., Ste. 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States of America
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/894,818B
; FILING DATE: 20-MAY-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP96/03253
; FILING DATE: 07-NOV-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 323285/1995
; FILING DATE: 12-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Browdy, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: TAKAKURA-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-5197
; TELEFAX: (202) 737-3528
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1566 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double

TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; FEATURE:
; OTHER INFORMATION: /note= N at position 1283 is G or T.
US-08-894-818B-4
Alignment Scores:
Pred. No.: 1.13e-32 Length: 1566
Score: 422.50 Matches: 138
Percent Similarity: 43.96% Conservative: 55
Best Local Similarity: 31.44% Mismatches: 167
Query Match: 18.67% Indels: 79
DB: Gaps: 17

US-09-985-689A-3 (1-433) x US-08-894-818B-4 (1-1566)
QY 16 AsnTyrGlyLeuTyrGlyGlnValValAlaValAlaAspThrGlyLeuAspThr 35
Db 58 AACITGGGATATGATGTTCTTGAATCAATAGGAATAATGACACTGGAATTGAC--- 114
QY 36 GlyArgAsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAla 55
Db 115 -----GCTTCTCATCCAGATCTCCAGGAAAAGTA----- 144
QY 56 LeuGlyArgThrAsnAsnAlaAsn-----AspProAsnGlyHisGly 69
Db 145 ATTGGGTGGGTAGATTTTGTCAATGCTAGGAGTTATCCATACCATGACCATGGACATGGA 204
QY 70 ThrHisValAlaGlySerValLeuGly-----AsnAlaLeuAsnLysGly 84
Db 205 ACTCATGTAGCTTCAATAGCAGCTGGTACTGGAGCAGCAAGTATATGCAAGTACACAGGGA 264
QY 85 MetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerSerGlyGly 102
Db 265 ATGGCTCCAGGAGCTAAGCTGCGGGAATTAAAGTTCTAGGTGCGGATGTTCTGGAGGC 324
QY 103 LeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaThrAsnAlaGlyAla 122
Db 325 ATATCTACTATAATAATTAAGGGAGTTGAGTGGGCCCTTGATACAAAGATAAGTACCGAAT 384
QY 123 ArgIleHisThrAsnSerTyrGlyAlaProValAsnGlyAlaTyrThrAlaAsnSerArg 142
Db 385 AAGGTCAATTATCTTCTTGGTTCAAGCCAGAGCTCAGATGTTACTGACGCTCTAAGT 444
QY 143 GlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAlaGlyAsnGlu 162
Db 445 CAGGCTGTTAATGCGAGCTGGGATGCTGGATTGTTGTGGTGGCGCTGGAAACAGT 504
QY 163 GlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGly 182
Db 505 GGACCTCAACAGTATACATCGTTCTCCAGCAGCTGCAAGCAAAAGTTATTACAGTTGGA 564
QY 183 AlaThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsnProAsnHisIleAla 202
Db 565 GCCGTTGACAGTAT-----GGAACCTAGCATCGGTCAACCA 591
QY 203 GlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspValThrAlaPro 222
Db 592 AGCTTCTCAAGCAGAGGGCCCACTGACAGCGGAGGCTTAAGCCTGAGTTGTTGCTCCA 651
QY 223 GlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheThrPalaAsn 242
Db 652 CGAAACTCGATAATTGCTGCCAGAGCAAGT-----GGAACCTAGCATCGGTCAACCA 702
QY 243 TyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGly 262
Db 703 ATTAATGACTATTACAGCAGCTCTCTGGGACATCAATGGCACTCTCTCAGTAGCTGT 762
QY 263 AsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro----- 279
Db 763 ATTGCAGCCCTCTTGTCTCAA-----GCACACCGAGCTGGACTCCAGACAAAGTA 813
QY 280 LysProSerLeuIleLysAlaAlaLeuIleAlaGlyAlaThrAspValGly---LeuGly 298

```
Db 814 AAAACAGCCCTCATGAAACTGCTGATATCGTAAAGCCAGATGAATAGCGGATATAGCC 873
Qy 299 TyrProSerGlyAspGlnGlyTyrPheValThrLeuAspGlySerLeuAsnValala 318
Db 874 TAC-----GGTGCAGGTAGGTTTATGTCATACAGGCTATTAAC----- 912
Qy 319 TyrValAsnGluAlaThrAlaLeuThrThrGlyGlnLysAla-----Thr 333
Db 913 TACGATAACTATGCAAGCTAGTGTCTACTGGATATGTTGCCAACAAGGCGACCAACT 972
Qy 334 TyrSerPheGlnThrGlnAlaGlyLysProLeuLysIleSerLeuValTyrThrAspAla 353
Db 973 CACCAAGTTTGGTATTAGCGGAGCTTGGTTCGTAACCTGCCATATATCTGGGCAATGCC 1032
Qy 354 ProGlySerThrThrAlaSerTyrThrLeuValAsnAspLeuValIleThrAla 373
Db 1033 AAT-----AGCGACCTTGTATCTTTACCTCTACGAT 1062
Qy 374 ProAsnGlyGlnLysTyrValGlyAsnAspPheSerTyrProTyrAspAsnAsnTyrAsp 393
Db 1063 CCCAATGAAACCCAGGTT-----GACTACTCTTAC-----ACC 1095
Qy 394 GlyAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIle 413
Db 1096 GCCTACTATGGATTGAAAGGTTGGTTATTACACCCCACTGATGGACATGGACAAATT 1155
Qy 414 GluValGlnAlaTyrAsnValProSerGlyProGlnArgPheSerLeuAlaIleVal 432
Db 1156 AAGGTTTGAAGCTAC-----AGCGGAAGTGCAAACTATCAAGTAGATGTGTA 1203
```

RESULT 9

US-08-894-818B-34
Sequence 34, Application US/088994818B
Patent No. 6261822

GENERAL INFORMATION:

APPLICANT: TAKAKURA, Hikaru
APPLICANT: MORISHITA, Mio
APPLICANT: YAMAMOTO, Katsuhiko
APPLICANT: MITTA, Masatoshi
APPLICANT: ASADA, Kiyoko
APPLICANT: TSUNASAWA, Susumu
APPLICANT: KATO, Ikunoshin
TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Browdy and Neimark
STREET: 419 Seventh Street N.W., Ste. 300
CITY: Washington
STATE: D.C.
COUNTRY: United States of America
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/894,818B
FILING DATE: 20-MAY-1998
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP96/03253
FILING DATE: 07-NOV-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 323285/1995
FILING DATE: 12-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Browdy, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: TAKAKURA=1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-5197

TELEFAX: (202) 737-3528
INFORMATION FOR SEQ ID NO: 34:
SEQUENCE CHARACTERISTICS:
LENGTH: 1962 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
US-08-894-818B-34

Alignment Scores:

Pred. No.: 159e-32 Length: 1962
Score: 422.50 Matches: 138
Percent Similarity: 43.96% Conservative: 55
Best Local Similarity: 31.44% Mismatches: 167
Query Match: 18.67% Indels: 79
DB: 3 Gaps: 17

US-09-985-689A-3 (1-433) x US-08-894-818B-34 (1-1962)

```
Qy 16 AsnTyrGlyLeuTyrGlyGlnGlyGlnValAlaValAlaAspThrGlyLeuAspThr 35
Db 454 AACTGGGATATGATGTTCTGGAATCACAATAGGAATAATTGACACTCGAATTGAC--- 510
Qy 36 GlyArgAsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAla 55
Db 511 -----CCTTCTCATCCAGATCTCCAGAGGAAAGTA----- 540
Qy 56 LeuGlyArgThrAsnAsnAlaAsn-----AspProAsnGlyHisGly 69
Db 541 ATTGGGTGGGTAGATTGTCATATGTTAGTAGGAGTTATCCATACGATGACCATGCACATGA 600
Qy 70 ThrHisValAlaGlySerValLeuGly-----AsnAlaLeuAsnLysGly 84
Db 601 ACTCATGTAGCTTCAATAGCAGCTGTGTACTGGAGCAGCAAGTAATGGCAAGTACAGGGA 660
Qy 85 MetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerSerGlyGly 102
Db 661 ATGGCTCCAGGAGCTAAGCTGGCGGGAATTAAGGTTCTAGTGCCGATGGTCTTCTGGAAGC 720
Qy 103 LeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsnAlaGlyAla 122
Db 721 ATATCTACTATAATAAAGGAGGTTAGTGCGCGCTGATACCAAGATAAGTACGGAATT 780
Qy 123 ArgIleHisThrAsnSerTyrGlyAlaProValAsnGlyValAlaTyrThrAlaAsnSerArg 142
Db 781 AAGGTCATTAACTCTTCTCTGTTTCAAGCCAGAGCTCAGATGTTACTGACGCTCTAAGT 840
Qy 143 GlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaGlyAsnGlu 162
Db 841 CAGGCTGTTTAATCAGCGTGGGATGCTGGATTAGTTGTTGTTGCCCTGGAAACAGT 900
Qy 163 GlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGly 182
Db 901 GGACCTAACAGTATACATCGGTTCTCCAGCAGCTGCAAGCAAGTTATTACAGTTGGA 960
Qy 183 AlaThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsnProAsnHisIleAla 202
Db 961 GCGGTTGCACCAAGTAT-----GATGTTATAACA 987
Qy 203 GlnPheSerSerArgGlyAlaThrArgAspGlyValGlyIleLysProAspValThrAlaPro 222
Db 988 AGCTTCTCAAGCAGAGGGCCAACTGCAGAGCGGAGGCTTAAGCTGAGGTTGTTGTTCCA 1047
Qy 223 GlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsn 242
Db 1048 GGAAACTGGATAATTGCTGCAGAGCAAGT-----GGAACTAGCATGGGTGCTCAACCA 1098
Qy 243 TyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGly 262
Db 1099 ATTAATGACTATTACAGCAGCTCTGGGACATCAATGGCACTCTCTACGATGTTGTT 1158
Qy 263 AsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro----- 279
```


QY 374 ProAsnGlyGlnLysTyrValGlyAsnAspPheSerTyrProTyrAspAsnAsnTrpAsp 393
 Db 1459 CCCATGGAACACAGTT-----GACTACTTTTAC-----ACC 1491
 QY 394 GlyArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThrThrIle 413
 Db 1492 GCCTACTATGATTGCAAAAGGTTGTTATTACAAACCACTGATGGAAACATGGACAATT 1551
 QY 414 GluValGlnAlaTyrAsnValProSerGlyProGlnArgPheSerLeuAlaIleVal 432
 Db 1552 AAGTTGTAGCTAC-----ACCGAAGTGCMAAACTATCAAGTAGATGTGGTA 1599

RESULT 11

US-08-894-818B-6
 ; Sequence 6, Application US/08894818B
 ; Patent No. 6261822
 ; GENERAL INFORMATION:
 ; APPLICANT: TAKAKURA, Hikaru
 ; APPLICANT: MORISHITA, Mio
 ; APPLICANT: YAMAMOTO, Katsuhiko
 ; APPLICANT: MITTA, Masanori
 ; APPLICANT: ASADA, Kiyozo
 ; APPLICANT: TSUNASAWA, Susumu
 ; APPLICANT: KATO, Ikunoshin
 ; TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
 ; NUMBER OF SEQUENCES: 42
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESS: Browdy and Neimark
 ; STREET: 419 Seventh Street N.W., Ste. 300
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: United States of America
 ; ZIP: 20004

COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/894,818B
 ; FILING DATE: 20-MAY-1998
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/JP96/03253
 ; FILING DATE: 07-NOV-1996

PRIOR APPLICATION DATA: JP 323285/1995
 ; APPLICATION NUMBER: 12-DEC-1995
 ; FILING DATE: 12-DEC-1995
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Browdy, Roger L.
 ; REGISTRATION NUMBER: 25,618
 ; REFERENCE/DOCKET NUMBER: TAKAKURA=1
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (202) 628-5197
 ; TELEFAX: (202) 737-3528
 ; INFORMATION FOR SEQ ID NO: 6:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1977 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear

MOLECULE TYPE: Genomic DNA

US-08-894-818B-6

Alignment Scores:

| | | | |
|------------------------|----------|---------------|------|
| Pred. No.: | 2,03e-30 | Length: | 1977 |
| Score: | 401.50 | Matches: | 130 |
| Percent Similarity: | 43.98% | Conservative: | 63 |
| Best Local Similarity: | 29.61% | Mismatches: | 167 |
| Query Match: | 17.74% | Indels: | 79 |
| DB: | 3 | Gaps: | 17 |

US-09-985-689A-3 (1-433) x US-08-894-818B-6 (1-1977)

QY 16 AsnTyrGlyLeuTyrGlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThr 35
 Db 454 AACTGGGATATGATCGTTCTGGAATCAATAGGAATTAATTCACATGGAATTGAC---- 510
 QY 36 GlyArgAsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAla 55
 Db 511 -----GCTTCTCATCCAGATCTCCAAGGAAGAATA----- 540
 QY 56 LeuGlyArgThrAsnAsnAlaAsn-----AspProAsnGlyHisGly 69
 Db 541 ATTGGGTGGGTAGATTGTCATATGGTAGGAGTTATCCATACGATGACCATGGACATGGA 600
 QY 70 ThrHisValAlaGlySerValLeuGly-----AsnAlaLeuAsnLysGly 84
 Db 601 ACTCATGTAGCTCAATAGCAGCTGCTACTGGAGCAGCAAGTAATGGCAAGTACAAAGGA 660
 QY 85 MetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerSerGlyGly 102
 Db 661 ATGGCTCCAGGAGCTAAGCTGGCGGAATTAAGGTTCTAGGTGCCGATGGTTCTTGAAGC 720
 QY 103 LeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsnAlaGlyAla 122
 Db 721 ATATCTACTATAATTAAAGGAGTTGAGTGGCGGTGATAACAAGATAAAGTACGGAAT 780
 QY 123 ArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrAlaAsnSerArg 142
 Db 781 AAGGTCATTAAATCTTCTCTGTTCAAGCCAGAGCTCCGACGGAACCCAGCTCCCTCAGT 840
 QY 143 GlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAlaGlyAsnGlu 162
 Db 841 CAGCGCGTCAACACCGCTGGGACCGCGGTAGTAGTCTGGCTGCCCGCGCAACAGC 900
 QY 163 GlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGly 182
 Db 901 GGSCCGAACACCTACACCGCTCGCTCACCGCGCGCGAGCAAGGTCTAATACCGTCGT 960
 QY 183 AlaThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsnProAsnHisIleAla 202
 Db 961 GCA-----GTTGACAGCAACGACCAACATCGCC 987
 QY 203 GlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspValThrAlaPro 222
 Db 988 AGCTTCTCAGCAGGGGACCGACCGGGACGGAAGGCTCAGCGGAAAGTCTCGTGGCCCC 1047
 QY 223 GlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsn 242
 Db 1048 GCGTTGACATCATAGCCCCGCGCGCCAGC-----GGAACCAAGCATGGGACCCCG 1098
 QY 243 TyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGly 262
 Db 1099 ATAAACGACTACTACACCAAGGCTCTGGACCAAGATGCCACCGCACGTTTCGGGC 1158
 QY 263 AsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro----- 279
 Db 1159 GTTGGCGCGCTCATCTCCAG-----GCCACCGAGCTGGACCCCGACCAAGGTG 1209
 QY 280 LysProSerLeuIleLysAlaLeuIleAlaGlyAlaThrAspValGly---LeuGly 298
 Db 1210 AAGACCGCCTCATCGAGACCGCGACATAGTGGCCCCCAAGAGATACGGACATCGCC 1269
 QY 299 TyrProSerGlyAspGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAla 318
 Db 1270 TAC-----GGTGGGTAGGTGACGTCATCAAGGCC-----ATCAAG 1308
 QY 319 TyrValAsnGluAlaThrAlaLeuThrThrGlyGln-----LysAlaThr 333
 Db 1309 TAGCAGACTACCCCAAGCTCCTTCCACCGCTCGTCCCGCAGCAAGGAAGCGCCACC 1368
 QY 334 TyrSerPheGlnThrGlnAlaGlyLysProLeuLysIleSerLeuValTrpThrAspAla 353
 Db 1369 CACACCTTCGACGTGACGCGCGCCACCTTCGTCGTGACCGCCACCTCTACTGGGAC----- 1422

QY 354 ProGlySerThrThrAlaSerTyrThrLeuValAsnAspLeuAspLeuValThrAla 373
Db 1423 -----ACGGGCTCGAGCGCATCGACCTCTACCTCTACGAC 1458
QY 374 ProAsnGlyGlnLysTyrValGlyAsnAspPheSerTyrProTyrAspAsnAsnTyrAsp 393
Db 1459 CCCAAC-----GGGACGAGTTGACTACTCTACCCGCTACTAC----- 1500
QY 394 GlyArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIle 413
Db 1501 -----GGCTTCGAGAGGTCGCTACTACAACCCGACCCCGAACTCGAGCGTC 1551
QY 414 GluValGlnAlaTyrAsnValProSerGlyProGlnArgPheSerLeuAlaIleVal 432
Db 1552 AAGTCTCGACTAC-----AAGGGCGCGCGGAACTACCAAGTTCGACGTCGTC 1599
RESULT 12
US-09-000-016-3
; Sequence 3, Application US/09000016
; Patent No. 6143541
; GENERAL INFORMATION:
; APPLICANT: AKIRA ARISAWA et al.
; TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC
; TITLE OF INVENTION: HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-DIHYDROPYRIDINE DERIV
; TITLE OF INVENTION: ITS EXPRESSION PRODUCT
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
; STREET: 2033 K Street, N.W., #800
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/000,016
; FILING DATE: January 30, 1998
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-721-8200
; TELEFAX: 202-721-8250
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2539 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Streptomyces viridosporus
; STRAIN: A-914
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 338...2539
; IDENTIFICATION METHOD: E
US-09-000-016-3
Alignment Scores: 5,94e-25 Length: 2539
Pred. No.: 348.50 Matches: 125
Score:

Percent Similarity: 38.99% Conservative: 45
Best Local Similarity: 28.67% Mismatches: 149
Query Match: 15.40% Indels: 117
DB: 3 Gaps: 17
US-09-985-689A-3 (1-433) x US-09-000-016-3 (1-2539)
QY 18 GlyLeuTyrGlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArg 37
Db 998 GGCTACGACGCGCAAGGGCGTGAAGATCGCGTCTGGACACCGGTGTCGACAGCG--- 1054
QY 38 AsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGly 57
Db 1055 -----CATCGGACCTGAAGGCGGCGTACCGCGTCCAAGAACTTCACC 1099
QY 58 ArgThrAsnAsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeu 77
Db 1100 GCCGCGCGCGCGCGCGGACAGGTGGCGCACCGCACCGCGCTCGCTCGATCGCGGCG 1159
QY 78 GlyAsnAlaLeuAsn-----LysGlyMetAlaProGlnAlaAsnLeuVal 92
Db 1160 GGCACGGCGCGCCAGTCCAGGGCAAGTACAAAGGCGTGCACCGCGCGCGGATCTCTC 1219
QY 93 PheGlnSerIleMetAspSerSerGlyGlyLeuGlyLeuProSerAsnLeuAsnThr 112
Db 1220 AACGGCAAGTCTCGACGACTCC-----GGTTTCGCGCAGCTCCGCGCATC 1267
QY 113 LeuPheSerGlnAlaTyrAsnAla-----GlyAlaArgIleHisThrAsnSerTyrGly 130
Db 1268 CTCGCGCGCATGAGTGGCGCGCGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1327
QY 131 Ala-----ProValAsnGlyAlaTyrThrAlaAsnSerArgGln 143
Db 1328 GGCATGGACACACCGGAGACCGACCGCTGGAGCGCGCG----- 1366
QY 144 ValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAlaGlyAsnGluGly 163
Db 1367 GTCGACAGCTGTCCGCGCGAGAGGGGTCTGTCTGCCATCGCGCGCGCGCAAGGGC 1426
QY 164 ProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAla 183
Db 1427 CCGGAGTCTG-----ATCGGTTCGCGCGCAGCGCGCGCGCGCGCGCGCGCGCG 1480
QY 184 ThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsnProAsnHisIleAlaGln 203
Db 1481 -----GTCGACGACAGGACAGCTCGCGGAC 1507
QY 204 PheSerSerArgGly---AlaThrArgAspGlyArgIleLysProAspValThrAlaPro 222
Db 1508 TTCTCTCTCACCGCGCGCGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1567
QY 223 GlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTyrAlaAsn 242
Db 1568 GCGGTGGACATCACGCGCGCTCGCGGAGGGCGACGACATCGCGCGCGCGCGCGCG 1627
QY 243 TyrAsnSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIleValAlaGly 262
Db 1628 GGACCGCGCGGCTACATGACCATCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1687
QY 263 AsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrProLysProSer 282
Db 1688 GCGCGCGCGCTCTCTGAGCAGCAG-----CACCCCGAC 1720
QY 283 LeuIleLysAlaAlaLeu-----IleAlaGlyAlaThrAspValGlyLeuGlyTyr 299
Db 1721 TGGACCTCCGCGCAACTGAGGGCGGCTCACCGGCTCCACCAAGGGCGGC---AAGTAC 1777
QY 300 ProSerGlyAspGlnGlyTyrPglYArgValThrLeuAspLysSerLeuAsnValAlaTyr 319
Db 1778 ACCCGGTTTCGACGAGGTTTCGCGCGGATCCAGCGCGCGCGCGCGCGCGCGCGCG 1828
QY 320 ValAsnGluAlaThrAlaLeuThrThrGlyGlnLysAlaThrTyrSerPheGlnThrGln 339

Db 1829 -----CAGACCGTG 1837
Qy 340 AlaGlyLysProLeuLysIleSerLeu-----ValTrp-----ThrAspAla 353
Db 1838 ATCGCGCAGCGGTTCGGTGGCTCCAGCAGTGGCGGCACACCGACGACGAG 1897
Qy 354 ProGlySerThrAlaSerThrLeuValAsnAspLeuValValle----- 371
Db 1898 CCGGTCCACAGCAGCTACCTACCGAACTTCGGCACCAGGACGTACCGCTGAAGCTG 1957
Qy 372 -----ThrAlaProAsnGlyGlnLysTyrValGlyAsnAspPheSerTyrPro 387
Db 1958 ACGTCGACCGCCAGCCAGCCCAAGGCAAG----- 1987
Qy 388 TyrAspAsnAsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAsnAlaProGln 407
Db 1988 -----GCGGCCCG 1996
Qy 408 SerGlyThrTyrThrIleGluValGlnAlaTyrAsnValProSerGly 423
Db 1997 GCGGCTCTTCACGCTGGCGCCACCACCGGTGACCGTCCCGCGGGC 2044

RESULT 13
US-09-514-340-3
Sequence 3, Application US/09514340
Patent No. 6361987
GENERAL INFORMATION:
APPLICANT: Akira ARISAWA et al.
TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC
HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-DIHYDROPYRIDINE DE
ITS EXPRESSION PRODUCT
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESS: Wenderoth, Lind & Ponack, L.L.P.
STREET: 2033 K Street, N.W., #800
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20006
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/514,340
FILING DATE: 28-Feb-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/000,016
FILING DATE: January 30, 1998
ATTORNEY/AGENT INFORMATION:
NAME: Lee Cheng
REGISTRATION NUMBER: 40,949
REFERENCE/DOCKET NUMBER: <Unknown>
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-721-8200
TELEFAX: 202-721-8250
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2539 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
ORIGINAL SOURCE:
ORGANISM: Streptomyces viridosporus
STRAIN: A-914
FEATURE:
NAME/KEY: CDS
LOCATION: 338...2539
IDENTIFICATION METHOD: E

SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-514-340-3
Alignment Scores:
Pred. No.: 5,94e-25 Length: 2539
Score: 348.50 Matches: 125
Percent Similarity: 38.99% Conservative: 45
Best Local Similarity: 28.67% Mismatches: 119
Query Match: 15.40% Indels: 17
Gaps: 4
US-09-985-689A-3 (1-433) x US-09-514-340-3 (1-2539)
Qy 18 GlyLeuTyrGlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArg 37
Db 998 GGTACAGCGCGAGGCGGTGAAGATCGCGCTCCGGACACCGGTGTGCACACGAGC--- 1054
Qy 38 AsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGly 57
Db 1055 -----CATCGGACCTGAAGGCGCGGTGACCGGTGCCAAGAACTTCACC 1099
Qy 58 ArgThrAsnAsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeu 77
Db 1100 GCGCGCGCGCGCGCGCGGACAAAGTGGGCGCACCGCACCTCGCTCGCTCGCGGGC 1159
Qy 78 GlyAsnAlaLeuAsn-----LysGlyMetAlaProGlnAlaAsnLeuVal 92
Db 1160 GGCACGCGCGCGCGCGCGGCAAGTACAAGGCGTCCGACCGCGCGCGCATCTC 1219
Qy 93 PheGlnSerIleMetAspSerSerGlyGlyLeuGlyLeuProSerAsnLeuAsnThr 112
Db 1220 ACGGCAAGTCTCTGACGACTCC-----GGTTTCGGCGACCATCCGCGCATC 1267
Qy 113 LeuPheSerGlnAlaTrpAsnAla-----GlyAlaArgIleHisThrAsnSerTrpGly 130
Db 1268 CTGCGCGCATGGAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1327
Qy 131 Ala-----ProValAsnGlyAlaTyrThrAlaAsnSerArgGln 143
Db 1328 GGCATGACACACCGGAGACCGACCGCTGGAGGGCGCG----- 1366
Qy 144 ValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAlaGlyAsnGluGly 163
Db 1367 GTCGACAAAGCTGTCCGCGGAGAGGCGTCTGTTCCGCGCATCGCGCGCGCGCGCGCG 1426
Qy 164 ProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAla 183
Db 1427 CCGGAGTCG-----ATCGGTTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1480
Qy 184 ThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsnProAsnHisIleAlaGln 203
Db 1481 -----GTCGACGACAAAGGACGCTCGCGCGAC 1507
Qy 204 PheSerSerArgGly---AlaThrArgAspGlyArgIleLysProAspValThrAlaPro 222
Db 1508 TTCTCTCCACCGCGCGCGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1567
Qy 223 GlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsn 242
Db 1568 GCGGTGACATCACCGCGCGCGCTCGCGGAGGCGGACACATCGCGCGAGGTCGCTGAG 1627
Qy 243 TyrAsnSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIleValAlaGly 262
Db 1628 GGACCGCGCGGTACATGACCATCTCCGCGCACGTGATGCGACCGCGCGCGCGCGCGCG 1687
Qy 263 AsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrProLysProSer 282
Db 1688 GCGCGCGCGCTCTGAAGCAGCAG-----CACCCCGAC 1720
Qy 283 LeuIleLysAlaAlaLeu-----IleAlaGlyAlaThrAspValGlyLeuGlyTyr 299
Db 1721 TGGACCTCCGCGCAACTGAAGCGCGCGCTCACCGGCTCCACCAAGGCGCGC---AAGTAC 1777


```
QY 300 ProSerGlyAspGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 319
Db 1778 ACCCGCTTCAGCAGCGGTTCCGGCGGATCCAGCGCGACAGCGCTCCAG-----1828
QY 320 ValAsnGluAlaThrAlaLeuThrThrGlyGlnLysAlaThrTyrSerPheGlnThrGln 339
Db 1829 -----CAGACCGTG 1837
QY 340 AlaGlyLysProLeuLysLeuSerLeu-----ValTrp-----ThrAspAla 353
Db 1838 ATCGCGCGCGCGCTCGGTGAGCTTCGGCGTCCAGCTGCGCGCACCGACCGACGAG 1897
QY 354 ProGlySerThrThrAlaSerTyrThrLeuValAsnAspLeuValIle-----371
Db 1898 CCGGTACCAAGCAGCTGACCTACCGCAACCTCGGCACCCAGGACGTACAGCTGAAGCTG 1957
QY 372 -----ThrAlaProLengGlnLysTyrValGlyAsnAspPheSerTyrPro 387
Db 1958 ACGTGACCGCGCCACCGACCGGCAAG-----1987
QY 388 TyrAspAsnAsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAsnAlaProGln 407
Db 1988 -----CGGCGCCCG 1996
QY 408 SerGlyThrTyrThrLeuValGlnAlaTyrAsnValProSerGly 423
Db 1997 GCGGGCTTCTTCAGCTGGCGCGCACCGGTGACCGTCCCGCGCGGC 2044

RESULT 14
US-09-000-016-1
; Sequence 1, Application US/09000016
; Patent No. 6143541
; GENERAL INFORMATION:
; APPLICANT: Akira ARISAWA et al.
; TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC
; TITLE OF INVENTION: HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-DIHYDROXYRIBINE DERIV
; TITLE OF INVENTION: ITS EXPRESSION PRODUCT
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
; STREET: 2033 K Street, N.W., #800
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/000,016
; FILING DATE: January 30, 1998
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-721-8200
; TELEFAX: 202-721-8250
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2809 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; ORIGINAL SOURCE:
```

```
ORGANISM: Streptomyces viridosporus
STRAIN: A-914
ORIGINAL SOURCE:
ORGANISM: Streptomyces antibioticus
STRAIN:
FEATURE:
NAME/KEY: CDS
LOCATION: 338...2539
IDENTIFICATION METHOD: E
FEATURE:
NAME/KEY: CDS
LOCATION: 2540...2809
IDENTIFICATION METHOD: P
US-09-000-016-1
Alignment Scores:
Pred. No.: 2,19e-24 Length: 2809
Score: 343.50 Matches: 124
Percent Similarity: 38.76% Conservative: 45
Best Local Similarity: 28.44% Mismatches: 150
Query Match: 15.18% Indels: 117
DB: 3 Gaps: 17
US-09-985-689A-3 (1-433) x US-09-000-016-1 (1-2809)
QY 18 GlyLeuTyrGlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArg 37
Db 998 GGCTACGACGCGCAAGGCGGTGAAGATCGCGTCTGGACACCGGTGTCGACAGC---1054
QY 38 AsnAspSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGly 57
Db 1055 -----CATCGGACCTGAGGCGGGGTGACCGGTCCAAAGAACTTCACC 1099
QY 58 ArgThrAsnAsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeu 77
Db 1100 GCCCGCGCGCGCGCGCGGACAGGTGCGGCACCGCACCGCTCGCTCGATCGCGCG 1159
QY 78 GlyAsnAlaLeuAsn-----LysGlyMetAlaProGlnAlaAsnLeuVal 92
Db 1160 GGCACGGCGCGCGCGCGCGCGTCCAGGCGCAAGTACAGGGCGTGCACCGCGCGCGATCTC 1219
QY 93 PheGlnSerIleMetAspSerSerGlyLeuGlyLeuGlyLeuProSerAsnLeuAsnThr 112
Db 1220 AACGCAAGGTCTCGACGACTCC-----GGTTTCGCGCGACGACTCCGGCATC 1267
QY 113 LeuPheSerGlnAlaTrpAsnAla-----GlyAlaArgIleHisThrAsnSerTrpGly 130
Db 1268 CTCGCGGCGATGGAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1327
QY 131 Ala-----ProValAsnGlyAlaTyrThrAlaAsnSerArgGln 143
Db 1328 GGCATGGACACACCGGAGACCGCGCTGGAGCGCGCG-----1366
QY 144 ValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaGlyAsnGluGly 163
Db 1367 GTCGCAAGCTGTCCGCCGAGAGGGGTCTGTTCGCCATCGCGCGCGCGCAACGAGGCG 1426
QY 164 ProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAla 183
Db 1427 CCGGAGTCG-----ATCGGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1480
QY 184 ThrGluAsnTyrArgProSerPheIleAlaAspAsnProAsnHisIleAlaGln 203
Db 1481 -----GTCGACGACAGGACCACTGCCGAC 1507
QY 204 PheSerSerArgGly-----AlaThrArgAspGlyArgIleLysProAspValThrAlaPro 222
Db 1508 TTCTCTCCACCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTCC 1567
QY 223 GlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsn 242
Db 1568 GCGGTGGACATCACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTGG 1627
```

QY 243 TyrAsnSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIleValalaGly 262
Db 1628 GGACGGCGCGGTACATGACCATCTCCGGCAGCTGATGGCAGCCCGCAGCTGCGGGC 1687
QY 263 AsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyLeuThrProLysProSer 282
Db 1688 GGGCGGCGCTCTGAGCAGCAG-----CACCGCGAC 1720
QY 283 LeuIleLysAlaAlaLeu-----IleAlaGlyAlaThrAspValGlyLeuGlyTyr 299
Db 1721 TGGACCTCGCGGAATGAAGCGCGCTCACCGCTCCACCAAGGGCGGC---AAGTAC 1777
QY 300 ProSerGlyAspGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 319
Db 1778 ACCCGTTCCAGCAGGGTTCCGGCGGATCCAGCGCGACAGGCGCTCAG----- 1828
QY 320 ValAsnGluAlaThrAlaLeuThrThrGlyGlnLysAlaThrTyrSerPheGlnThrGln 339
Db 1829 -----CAGACCGTG 1837
QY 340 AlaGlyLysProLeuLysIleSerLeu-----ValTyr-----ThrAspAla 353
Db 1838 ATCGCGACCGGCTCGGTGAGCTTCCGGCTCCAGCAGTGGCCGACCGACCGACGACGAG 1897
QY 354 ProGlySerThrThrAlaSerTyrThrLeuValAsnAspLeuAspLeuValIle----- 371
Db 1898 CCGGTACCAAGCAGCTGACCTACCGCAACCTCGGCACCGCAGCTGAGCTGAGCTG 1957
QY 372 -----ThrAlaProAsnGlyGlnLysTyrValGlyAsnAspPheSerTyrPro 387
Db 1958 ACCTGACCGCCACCGACCCCAAGGCAAG----- 1987
QY 388 TyrAspAsnAsnTyrAspGlyArgAsnAsnValGluAsnValPheIleAsnAlaProGln 407
Db 1988 -----CGCGCCCG 1996
QY 408 SerGlyThrTyrThrIleGluValGlnAlaTyrAsnValProSerGly 423
Db 1997 GCGGGCTTCTTCAGCTGGCGGCCACCGCGTGACCGTCCCGCGCGGC 2044

RESULT 15
US-09-514-340-1
Sequence 1, Application US/09514340
Patent No. 6361987
GENERAL INFORMATION:
APPLICANT: AKIRA ARISAWA et al.
TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC
HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-DIHYDROPYRIDINE DE
ITS EXPRESSION PRODUCT
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
STREET: 2033 K Street, N.W., #800
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20006
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/514,340
FILING DATE: 28-Feb-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/000,016
FILING DATE: January 30, 1998
ATTORNEY/AGENT INFORMATION:
NAME: Lee Cheng
REGISTRATION NUMBER: 40,949
REFERENCE/DOCKET NUMBER: <Unknown>

TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-721-8200
TELEFAX: 202-721-8250
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2809 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
ORIGINAL SOURCE:
ORGANISM: Streptomyces antibioticus
STRAIN: <unknown>
FEATURE:
NAME/KEY: CDS
LOCATION: 338...2539
IDENTIFICATION METHOD: E
FEATURE:
NAME/KEY: CDS
LOCATION: 2540...2809
IDENTIFICATION METHOD: P
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-514-340-1
Alignment Scores:
Pred. No.: 2,19e-24 Length: 2809
Score: 343.50 Matches: 124
Percent Similarity: 38.76% Conservative: 45
Best Local Similarity: 28.44% Mismatches: 150
Query Match: 15.18% Indels: 117
Gaps: 17
US-09-985-689A-3 (1-433) x US-09-514-340-1 (1-2809)
QY 18 GlyLeuTyrGlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArg 37
Db 998 GGCTACGACGCAAGGCGGTGAAGATCGCGCTCTGGACACCGGTGTGCACGAGC--- 1054
QY 38 AsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGly 57
Db 1055 -----CATCGGACCTGAAGGGCGGTGACCGCTCCCAAGAACTTCACC 1099
QY 58 ArgThrAsnAsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeu 77
Db 1100 GCG 1159
QY 78 GlyAsnAlaLeuAsn-----LysGlyMetAlaProGlnAlaAsnLeuVal 92
Db 1160 GGCACGGCTC 1219
QY 93 PheGlnSerIleMetAspSerSerGlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThr 112
Db 1220 AACGGCAAGGTCTCTCGACGACTCC-----GGTTTCGGCGACGACTCCGGCATC 1267
QY 113 LeuPheSerGlnAlaTyrAsnAla-----GlyAlaArgIleHisThrAsnSerTyrGly 130
Db 1268 CTGCGCGCATGGAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1327
QY 131 Ala-----ProValAsnGlyAlaTyrThrAlaAsnSerArgGln 143
Db 1328 GGCATGGACACACCGGAGACCGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1366
QY 144 ValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAlaGlyAsnGly 163
Db 1367 GTGACAAAGCTGTCCGCGGAGAGGGGCTCTTTCGCCATCGCGCGCGCGCGCGCGCG 1426
QY 164 ProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAla 183
Db 1427 CCGAGATCG-----ATCGGTTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1480
QY 184 ThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsnProAsnHisIleAlaGln 203

```

Db 1481 -----GTCGACGACAGCAAGCTGCCGAC 1507
Qy 204 PheSerSerArggly---AlaThrArgAspGlyArgIleLeuProAspValThrAlaPro 222
Db 1508 TTCTCTCCACGGCCCGCTCCGCGAGCGGGCCATCAAGCCGACGTACCGCTCCC 1567
Qy 223 GlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsn 242
Db 1568 GCGGTGGACATCACGGCCCGCTCCGCGAGGGCAACGACATCGGCCAGGAGTCCGGTGAG 1627
Qy 243 TyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGly 262
Db 1628 GGCACCGCGCGGTACATGACCATCTCCGCGACGTGATCGCGACCCCGACGTCCGGGCGC 1687
Qy 263 AsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrProLysProSer 282
Db 1688 GCGCGCGCCCTCTGAAGCAGCAG-----CACCCCGAC 1720
Qy 283 LeuIleLysAlaAlaLeu-----IleAlaGlyAlaThrAspValGlyLeuGlyTyr 299
Db 1721 TGACCTCCGCCCACTGAAGGGCGGCTCACCGGTCCACCAAGGGCGGC---AAGTAC 1777
Qy 300 ProSerGlyAspGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 319
Db 1778 ACCCGGTTCGAGCAGGCTTCGGCGCGATCCAGCGCGACAAAGCGCTCCAG----- 1828
Qy 320 ValAsnGluAlaThrAlaLeuThrThrGlyGlnLysAlaThrTyrSerPheGlnThrGln 339
Db 1829 -----CAGACCGTG 1837
Qy 340 AlaGlyLysProLeuLysIleSerLeu-----ValTrp-----ThrAspAla 353
Db 1838 ATCGCGACCCGGTCTCGGTGAGTTCGGCGTCCAGCAGTGGCGGCACACCGACGACGAG 1897
Qy 354 ProGlySerThrThrAlaSerTyrThrLeuValAsnAspLeuAspLeuValIle----- 371
Db 1898 CCGGTCCACCAAGCGTGTACCTCCGCACTCCGACCCCGACCGACGTCCGCTGAAGCTG 1957
Qy 372 -----ThrAlaProAsnGlyGlnLysTyrValGlyAsnAspPheSerTyrPro 387
Db 1958 ACGTCGACCGGCACCGACCCCAAGGGCAAG----- 1987
Qy 388 TyrAspAsnAsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAsnAlaProGln 407
Db 1988 -----GCGGCCCG 1996
Qy 408 SerGlyThrTyrThrIleGluValGlnAlaTyrAsnValProSerGly 423
Db 1997 GCGGGCTTCTTCACGCTGGCGGCCACCGCGTGACCGTCCCGCGGGGC 2044

```

Search completed: April 4, 2004, 12:02:03
Job time : 90.4512 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 4, 2004, 08:10:21 ; Search time 327.248 Seconds
(without alignments)
4948.852 Million cell updates/sec

Title: US-09-985-689A-3
Perfect score: 2263
Sequence: 1 NDVARGIVKADVAQNNYGLY.....EVQAYNVPSGPQRFSLAIVH 433

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2456186 seqs, 1870095128 residues

Total number of hits satisfying chosen parameters: 4932372

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US0985689/runat_31032004_161809_4271/app_query.fasta_1.3498
-DB=Published Applications NA -GFM=fastap -SUFFIX=rnpb -MINWATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blcsum62
-TRANS=human40 cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODS=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US0985689@cgn_1_1_601@runat_31032004_161809_4271
-ICPU=3 -ICPU=3 -NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LUNGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-XGAPOP=6 -XGAPEXT=7 -XGAPOP=10 -XGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA.*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/1/pubpna/US09D_PUBCOMB.seq.*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq.*
16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq.*
17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq.*
18: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
19: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------|
|------------|-------|-------------|--------|----|-------------|

| | | | | | | |
|----|--------|------|---------|----|----------------------|-------------------|
| 1 | 1994.5 | 88.1 | 1305 | 15 | US-10-385-662-1 | Sequence 1, Appli |
| 2 | 443 | 19.6 | 1977 | 13 | US-10-090-624-11 | Sequence 11, Appl |
| 3 | 422.5 | 18.7 | 1236 | 13 | US-10-090-624-2 | Sequence 2, Appli |
| 4 | 422.5 | 18.7 | 1962 | 13 | US-10-090-624-15 | Sequence 15, Appl |
| 5 | 346.5 | 15.3 | 135638 | 14 | US-10-314-857-1 | Sequence 1, Appli |
| 6 | 336 | 14.8 | 1329 | 9 | US-09-974-300-1934 | Sequence 1934, Ap |
| 7 | 329 | 14.5 | 3624 | 14 | US-10-156-761-5701 | Sequence 5701, Ap |
| 8 | 329 | 14.5 | 9025608 | 14 | US-10-156-761-1 | Sequence 1, Appli |
| 9 | 327 | 14.4 | 3417 | 14 | US-10-156-761-3306 | Sequence 3306, Ap |
| 10 | 327 | 14.4 | 9025608 | 14 | US-10-156-761-1 | Sequence 1, Appli |
| 11 | 305 | 13.5 | 4765 | 13 | US-10-090-624-5 | Sequence 5, Appli |
| 12 | 296.5 | 13.1 | 3743 | 10 | US-09-927-827-29 | Sequence 29, Appl |
| 13 | 295 | 13.0 | 1560 | 15 | US-10-084-846A-113 | Sequence 113, App |
| 14 | 295 | 13.0 | 59816 | 15 | US-10-084-846A-2 | Sequence 2, Appli |
| 15 | 295 | 13.0 | 59816 | 15 | US-10-084-846A-2 | Sequence 2, Appli |
| 16 | 282.5 | 12.5 | 3303 | 14 | US-10-156-761-5384 | Sequence 5384, Ap |
| 17 | 281.5 | 12.4 | 2166 | 12 | US-10-344-231-17 | Sequence 17, Appl |
| 18 | 281.5 | 12.4 | 2166 | 12 | US-10-363-332A-17 | Sequence 17, Appl |
| 19 | 269 | 11.9 | 1140 | 8 | US-08-322-678-11 | Sequence 11, Appl |
| 20 | 269 | 11.9 | 1140 | 16 | US-10-323-324-11 | Sequence 11, Appl |
| 21 | 269 | 11.9 | 1143 | 14 | US-10-313-853-6 | Sequence 6, Appli |
| 22 | 269 | 11.9 | 2588 | 12 | US-10-344-231-20 | Sequence 20, Appl |
| 23 | 269 | 11.9 | 2588 | 12 | US-10-363-332A-20 | Sequence 20, Appl |
| 24 | 269 | 11.9 | 3788 | 10 | US-09-927-827-33 | Sequence 33, Appl |
| 25 | 268 | 11.8 | 1140 | 8 | US-08-322-678-12 | Sequence 12, Appl |
| 26 | 268 | 11.8 | 1140 | 16 | US-10-323-324-12 | Sequence 12, Appl |
| 27 | 260.5 | 11.5 | 1506 | 9 | US-09-966-921A-1 | Sequence 1, Appli |
| 28 | 260.5 | 11.5 | 1330 | 9 | US-09-966-921A-5 | Sequence 5, Appli |
| 29 | 259.5 | 11.5 | 840 | 14 | US-10-209-812-1 | Sequence 1, Appli |
| 30 | 258.5 | 11.4 | 2267 | 12 | US-10-344-231-18 | Sequence 18, Appl |
| 31 | 258.5 | 11.4 | 2267 | 12 | US-10-363-332A-18 | Sequence 18, Appl |
| 32 | 252.5 | 11.2 | 1971 | 9 | US-09-974-300-1935 | Sequence 1935, Ap |
| 33 | 252 | 11.1 | 1485 | 9 | US-09-974-300-1938 | Sequence 1938, Ap |
| 34 | 247.5 | 10.9 | 2192 | 12 | US-10-424-559-112429 | Sequence 112429, |
| 35 | 247 | 10.9 | 3884 | 10 | US-09-927-827-34 | Sequence 34, Appl |
| 36 | 242.5 | 10.7 | 3452 | 10 | US-09-927-827-30 | Sequence 30, Appl |
| 37 | 242 | 10.7 | 1140 | 15 | US-10-146-905A-9 | Sequence 9, Appli |
| 38 | 238 | 10.5 | 1140 | 15 | US-09-920-118-13 | Sequence 13, Appl |
| 39 | 232 | 10.3 | 1332 | 14 | US-10-156-761-5689 | Sequence 5689, Ap |
| 40 | 231.5 | 10.2 | 4338 | 9 | US-09-891-711-3 | Sequence 3, Appli |
| 41 | 231.5 | 10.2 | 4338 | 12 | US-10-342-887-953 | Sequence 953, App |
| 42 | 230 | 10.2 | 1575 | 12 | US-10-344-231-19 | Sequence 19, Appl |
| 43 | 230 | 10.2 | 1575 | 12 | US-10-363-332A-19 | Sequence 19, Appl |
| 44 | 228 | 10.1 | 4198 | 9 | US-09-891-711-5 | Sequence 5, Appli |
| 45 | 227.5 | 10.1 | 522 | 10 | US-09-824-893A-98 | Sequence 98, Appl |

ALIGNMENTS

RESULT 1
US-10-385-662-1
; Sequence 1, Application US/10385662
; Publication No. US2004002432A1
; GENERAL INFORMATION:
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SAITO, TSUYOSHI
; APPLICANT: SAITO, KAZUHIRO
; APPLICANT: SUMITOMO, NOBUYUKI
; APPLICANT: IZAWA, YOSHIFUMI
; APPLICANT: SAEKI, KATSUHIKA
; APPLICANT: KOBAYASHI, TOHRU
; APPLICANT: NOMURA, MASAFUMI
; TITLE OF INVENTION: Alkaline protease
; FILE REFERENCE: 234938US0
; CURRENT APPLICATION NUMBER: US/10385,662
; CURRENT FILING DATE: 2003-03-12
; PRIOR APPLICATION NUMBER: JP 2002-081428
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: JP 2002-165987
; PRIOR FILING DATE: 2002-06-06
; PRIOR APPLICATION NUMBER: JP 2002-304230
; PRIOR FILING DATE: 2002-10-18
; PRIOR APPLICATION NUMBER: JP 2002-304231

;; PRIOR FILING DATE: 2002-10-18
;; NUMBER OF SEQ ID NOS: 20
;; SOFTWARE: Patent in version 3.1
;; SEQ ID NO 1
;; LENGTH: 1305
;; TYPE: DNA
;; ORGANISM: Bacillus sp. KSM-KP43
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: (1)..(1305)
;; OTHER INFORMATION:
US-10-385-662-1

Alignment Scores:
Pred. No.: 1,13e-208 Length: 1305
Score: 1994.50 Matches: 380
Percent Similarity: 94.0% Conservatives: 28
Best Local Similarity: 87.56% Mismatches: 25
Query Match: 88.14% Indels: 1
DB: 15 Gaps: 1

US-09-985-689A-3 (1-433) x US-10-385-662-1 (1-1305)

QY 1 AsnAspValAlaArgGlyLeuValAlaAspValAlaGlnAsnAsnTyrGlyLeuTyr 20
Db 1 AATGATGTTGCGGTGGAAATGTCAAAGCGGATGTGGCTCAGACGAGCTACGGGTGTAT 60
QY 21 GlyGlnGlyValValAlaAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 61 GGCACAGGACAGATCGTACGGTGTCCGATACAGGCTTGATACAGGTGCAATGACAGT 120
QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 121 TCGATGCATGAACCTTCGCGGGAATAATTACTGCATTATATGCAATTTGGACGCGAAT 180
QY 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn--- 79
Db 181 AATGCCAATGATACGATGGTCAATGATGATGATGATGATGATGATGATGATGATGATGAT 240
QY 80 AlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 99
Db 241 TCCACTAATAAGGAATGGCGCTCAGGCGAATCTAGTCTTCCATCTATCATGTGATAGC 300
QY 100 SerGlyGlyLeuGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsn 119
Db 301 GGTGGGGACTTGGAGGACTACCTTGAATCTGCAAACTTATTCAGCCAGCATACAGT 360
QY 120 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrAla 139
Db 361 GCTGTGCCAGAAATTCATACAACTCTCTGGGAGCAGCAGTGAATGGGCTTACACACA 420
QY 140 AsnSerArgGlnValAspGluTyrValArgAsnAspMetThrValLeuPheAlaAla 159
Db 421 GATTCAGAAATGGGATGACTATGCGGCAAAATGATATGATGATGATGATGATGATGATGAT 480
QY 160 GlyAsnGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 179
Db 481 GGGAAATGAAGCAGCCAGCGGGAACCATCAGTGCACCGACGACGATCAAAATGCAATA 540
QY 180 ThrValGlyAlaThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsnProAsn 199
Db 541 ACATGCGAGCTACGGAACCTCCGCCAAGCTTTGGCTTATGCGGCAATATCAAC 600
QY 200 HisIleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspVal 219
Db 601 CATGTGGCAGATCTCTTCACGTGGACCGCAAAAGGATGGACGATCAAAACCGGATGTC 660
QY 220 ThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerPhe 239
Db 661 ATGGCAGCGGAGCTTCATACACTATCAGCAAGATCTTCTCTTGGACCGGATCTCTCTTC 720
QY 240 TrpAlaAsnTyrAsnSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIle 259

Db 721 TGGGCGAACCATGACAGTAATAATATGCATATGGGTGGAAACGTCCTGCTACCCGATC 780
QY 260 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro 279
Db 781 GTTCTGGAAACGTGGCACAGCTTCTGTAGCATTTTGTGAAAAACAGAGGCATCACACCA 840
QY 280 LysProSerIleLysAlaAlaLeuIleAlaGlyAlaThrAspValGlyLeuGlyTyr 299
Db 841 AGCCTTCTCTATTAAGCGGCACTGATTGCGGTGACAGTCACATCGGCTTGCTAC 900
QY 300 ProSerGlyAspGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 319
Db 901 CCGAACGGTAACCAAGATGGGACGAGTGACATGATGATTAATCCCTGACCGTGGCTAT 960
QY 320 ValAsnGluAlaThrAlaLeuThrThrGlyGlnLysAlaThrThrPheGlnThrGln 339
Db 961 GTGAACGAGTCCAGTTCTCTATCCACGACCAAAAGCGAGCTACTCGTTTACTGCTACT 1020
QY 340 AlaGlyLysProLeuLysIleSerLeuValTrpThrAspAlaProGlySerThrAla 359
Db 1021 GCCGGCAAGCCTTTGAAATCTCCCTGGTATGCTGTGATGCTGCTGCGGACCACTGCT 1080
QY 360 SerTyrThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyGlnLysTyr 379
Db 1081 TCCGTAACGCTTGTCAATGATCTGGACCTTGTCTATTACCGCTCCAAATGGCACACAGTAT 1140
QY 380 ValGlyAsnAspPheSerTyrProTyrAspAsnAsnTrpAspGlyArgAsnValGlu 399
Db 1141 GTAGGAATGACTTACTTCGCCATACATGATACTGGATGCGCGCAATAACGATAGAA 1200
QY 400 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 419
Db 1201 AATGATTTTATTAATGCACCAAAAGCGGACGATATACAAATTCAGGTACAGGCTTATAAC 1260
QY 420 ValProSerGlyProGlnArgPheSerLeuAlaIleValHis 433
Db 1261 GTACCGTGGACACAGACCTTCTGTTGGCAATTTGTGAAT 1302

RESULT 2
US-10-090-624-11
; Sequence 11, Application US/10090624
; Publication No. US20020132335A1
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA-6
; CURRENT APPLICATION NUMBER: US/10/090,624
; CURRENT FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: 09/445,472
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 11
; LENGTH: 1977
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-10-090-624-11

Alignment Scores:
Pred. No.: 2,82e-38 Length: 1977
Score: 443.00 Matches: 137
Percent Similarity: 46.03% Conservatives: 66
Best Local Similarity: 31.07% Mismatches: 170
Query Match: 19.58% Indels: 68
DB: 13 Gaps: 16

[illegible]

DB 1156 AGGGTTGAAGTAC-----AGCGAAGTGCAAACTATCAAGTAGATGTGGTA 1203

RESULT 4

US-10-090-624-15

Sequence 15, Application US/10090624

Publication No. US20020132335A1

GENERAL INFORMATION:

APPLICANT: TAKAKURA, Hikaru

APPLICANT: MORISHITA, Mio

APPLICANT: SHIMOJO, Tomoko

APPLICANT: ASADA, Kiyozo

APPLICANT: KATO, Ikunoshin

TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE

FILE REFERENCE: TAKAKURA=6

CURRENT APPLICATION NUMBER: US/10/090,624

CURRENT FILING DATE: 2002-03-06

PRIOR APPLICATION NUMBER: 09/445,472

PRIOR FILING DATE: 1999-12-06

PRIOR APPLICATION NUMBER: 151969/1997

PRIOR FILING DATE: 1997-06-10

NUMBER OF SEQ ID NOS: 33

SOFTWARE: Patentin version 3.0

SEQ ID NO 15

LENGTH: 1962

TYPE: DNA

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Synthetic

US-10-090-624-15

Alignment Scores:

Pred. No.: 4,93e-36 Length: 1962

Score: 422.50 Matches: 138

Percent Similarity: 43.96% Conservative: 55

Best Local Similarity: 31.44% Mismatches: 167

Query Match: 18.67% Indels: 79

DB: 13 Gaps: 17

US-09-985-689A-3 (1-433) x US-10-090-624-15 (1-1962)

| | | | |
|----|-----|---|-----|
| QY | 16 | AsnTyrGlyLeuTyrGlyGlyGlnValValAlaAspThrGlyLeuAspThr | 35 |
| DB | 454 | AAC TTGGGATATGATGTTCTTGGAAATCACAATAGGAATAATTGACCTGGAATTGC | 510 |
| QY | 36 | GlyArgAsnAspSerSerMetHisGluAlaPheArgGlyValIleThrAlaLeuTyrAla | 55 |
| DB | 511 | -----GOTTCTATCCAGATCTCCAGGAAGAATGTA----- | 540 |
| QY | 56 | LeuGlyArgThrAsnAsnAlaAsn-----AspProAsnGlyHisGly | 69 |
| DB | 541 | ATTGGGTGGGTAGATTTTGTCAATGGTAGGAGTTATCCATACGATGACCATGGACATGGA | 600 |
| QY | 70 | ThrHisValAlaGlySerValLeuGly-----AsnAlaLeuAsnLysGly | 84 |
| DB | 601 | ACTCATGTAGTTC AATAGCAGCTGGTACTGGAGCAGCAAGTAATGGCAAGTACAAAGGA | 660 |
| QY | 85 | MetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerSerGlyGly | 102 |
| DB | 661 | ATGGCTCCAGAGCTAAGCTGGCGGAATTAAGGTTCTAGTGGCCGATGGTTCTGGAAGC | 720 |
| QY | 103 | LeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsnAlaGlyAla | 122 |
| DB | 721 | ATATCTACTATAATTAAGGGAGTTGAGTGGGCCGCTTGATATCAAAAGATAAGTACGAATT | 780 |
| QY | 123 | ArgIleHisThrAsnSerTyrGlyAlaProValAsnGlyAlaTyrThrAlaAsnSerArg | 142 |
| DB | 781 | AAGGTCATTAATCTTCTTCTGGTTCAAGCCAGAGCTCAGATGGTACTGACGCTCTAAGT | 840 |
| QY | 143 | GlnValAspGluTyrValIArgAsnAsnAspMetThrValLeuPheAlaAlaGlyAsnGlu | 162 |
| DB | 841 | CAGGCTGTAAATGCAGCGTGGGATCTGGATTAGTTGTTGTTGGTGGCCGCTGGAACACT | 900 |
| QY | 163 | GlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGly | 182 |

```
Db 901 GGACCTAACAGTATACATCGTTCTCCAGCAGCTGCAAGCAAAAGTTATACAGTTGGA 960
Qy 183 AlaThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsnProAsnHisIleAla 202
Db 961 GCCGTTGACAAGTAT-----GATGTTATAACA 987
Qy 203 GlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspValThrAlaPro 222
Db 988 AGCTTCTCAACAGAGGCCCAACTGCACAGCGCAGGCTTAAGCCTGAGGTGTGTGCTCCA 1047
Qy 223 GlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTyrAlaAsn 242
Db 1048 GGAACCTGGATTAATGTCGCCAGGCAAGT-----GGAACCTAGCATGGGTCAACCA 1098
Qy 243 TyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGly 262
Db 1099 ATTAATGACTATTACACAGCAGCTCTCGGACATCAATGGCACTCCTCAGTAGCTGGT 1158
Qy 263 AsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro----- 279
Db 1159 ATTGCAGCCCTCTTGCTCCAA-----GCACACCGAGCTGGATCCAGCAAAAGTA 1209
Qy 280 LysProSerLeuIleLysAlaAlaLeuIleAlaGlyAlaThrAspValGly---LeuGly 298
Db 1210 AAAACAGCCCTCATAGAAACTGCTGATATCGTAAAGCCAGATGAATAGCGGATATAGCC 1269
Qy 299 TyrProSerGlyAspGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAla 318
Db 1270 TAC-----GGTGCAGGTAGGTTTAAATGCATACAAAGGCTATAAACC----- 1308
Qy 319 TyrValAsnGluAlaThrAlaLeuThrThrGlyGlnLysAla-----Thr 333
Db 1309 TACGATAACTATGCAAGCTAGTGTCTACTGGATATGTTGCCAACAAAGGCGACCAAACT 1368
Qy 334 TyrSerPheGlnThrGlnAlaGlyLysProLeuLysIleSerLeuValThrThrAspAla 353
Db 1369 CACCAAGTTCGTTATTAGCGGAGCTCGTTCGTAACCTGCACATTATCTGGGCAAAATGCC 1428
Qy 354 ProGlySerThrThrAlaSerTyrThrLeuValAsnAspLeuAspLeuValIleThrAla 373
Db 1429 AAT-----AGCGACCTTGATCTTACCTCTACGAT 1458
Qy 374 ProAsnGlyGlnLysTyrValGlyAsnAspPheSerTyrProTyrAspAsnAsnTrpAsp 393
Db 1459 CCCAATGGAAACCCAGGTT-----GACTACTCTTAC-----ACC 1491
Qy 394 GlyArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIle 413
Db 1492 GCCTACTATGNTTCGAAAGGTTGGTTATTACAAACCACTGATGGNACATGGACAAATT 1551
Qy 414 GluValGlnAlaTyrAsnValProSerGlyProGlnArgPheSerLeuAlaIleVal 432
Db 1552 AAGGTTGTAAGCTAC-----AGCGGAAGTGCAAACTATCAAGTAGATGTGGTA 1599

RESULT 5
US-10-314-657-1
; Sequence 1, Application US/10314657
; Publication No. US20030175888A1
; GENERAL INFORMATION:
; APPLICANT: SHEN, Ben
; APPLICANT: TANG, Yi-Qiang
; APPLICANT: CHENG, Li
; TITLE OF INVENTION: Discrete Acyltransferases Associated with Type I Polyketide
; TITLE OF INVENTION: Syntheses and Methods of Use
; FILE REFERENCE: 054030-0021
; CURRENT APPLICATION NUMBER: US/10/314,657
; PRIOR FILING DATE: 2002-12-09
; PRIOR FILING DATE: PCT/US02/08937
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: US 60/278,935
; PRIOR FILING DATE: 2001-03-26
; NUMBER OF SEQ ID NOS: 214
```

```
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 135638
; TYPE: DNA
; ORGANISM: Streptomyces atroolivaceus
US-10-314-657-1

Alignment Scores:
Pred. No.: 4,928-25 Length: 135638
Score: 346.50 Matches: 140
Percent Similarity: 42.39% Conservative: 55
Best Local Similarity: 30.43% Mismatches: 154
Query Match: 15.31% Indels: 111
DB: 14 Gaps: 23

US-09-985-689A-3 (1-433) x US-10-314-657-1 (1-135638)
Qy 6 GlyIleValLysAlaAspValAlaGlnAsnAsn----- 16
Db 8067 GGCAGGGTGAAGGCGATCTGCCGACTCCACCGCCAGATCGGGCGCGCAGAGGTATGG 8126
Qy 17 ---TyrGlyLeuTyrGlyGlnGlyValValAlaValAlaAspThrGlyLeuAspThr 35
Db 8127 GCGGAGGGCGACACCGCGCAGGACGTGAAGGTGCGGATGCTGCACAGCGCGCGACACC 8186
Qy 36 GlyArgAsnAspSerSerMetHisGluAlaPheArgGlyLysIle-----ThrAlaLeu 53
Db 8187 -----GAACACCGGACCTGTGCGGCGAGGTGTCGCACAGCGCCAGC 8228
Qy 54 TyrAlaLeuGlyArgThrAsnAsnAlaAsnAspProAsnGlyHisGlyThrHisValAla 73
Db 8229 TTCGTCCCGCGGAGCAGCATCGCC---GACTACAAACGCGCAGCGCAGCATCGCC 8285
Qy 74 GlySerValLeuGly-----AsnAlaLeuAsnLysGlyMetAlaProGln 88
Db 8286 TCGACCATCTCGGCACCGCGCGCGCTCCGACGCAAGGAGCGGGGTGTCGCTCCGCG 8345
Qy 89 AlaAsnLeuValPheGlnSerIleMetAspSerSerGlyGlyLeuGlyGlyLeuProSer 108
Db 8346 GCGCGCTCTCGCTCGGCAAGGTGCTCACTCCGAGGCG-----AGCGCGCAGGAATCG 8399
Qy 109 AsnLeuAsnThrLeuPheSerGlnAlaTyrAsnAla-----GlyAlaArgIleHis 125
Db 8400 TGGATC-----ATCGCGGCGATGGATGGGCGCGCCGCGCAGCAAGCCAGGATCATC 8453
Qy 126 ThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrAlaAsn-----SerArg 142
Db 8454 AGCATGAGCTGGCG-----GGCGCGGTGCACACAGACGACCCGATGAGCCAG 8501
Qy 143 GlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAlaGlyAsnGlu 162
Db 8502 GCGTCGACGAACTACGCCACGACACGCGCGGTGTGTGTCGTGATCGCGCGGCGAACCGC 8561
Qy 163 GlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGly 182
Db 8562 GCGCGCACTCC-----ATCAGACGCGCGGTGCGGACACTCGCGGTGACCGTCGCGC 8615
Qy 183 AlaThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsnProAsnHisIleAla 202
Db 8616 GCC-----GTGCACTCCACCGACAGCTCGCC 8642
Qy 203 GlnPheSerSerArgGly---AlaThrArgAspGlyArgIleLysProAspValThrAla 221
Db 8643 GACTTCTCAGCAGCGCGCGCGTGCAGCGCGCGGCTGAAGCGGAGATCACCAGC 8702
Qy 222 ProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAla 241
Db 8703 CCGCGGCTCGACATCGTCGCGCGCTCGCTCAAGCGCGGCTCGCGGTAC----- 8756
Qy 242 AsnTyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleValAla 261
Db 8757 -----TACACCAGATGACGCGCACGTCTGATGGCGGACGCGCGCATCGCC 8801
```


QY 262 GlyAsnValAlaGluHisPheileLeuAsnArgGlyLeuThrProLysPro 281
 DB 8802 GGTGTCCGCGCGCTCTCCCGCGAGACACCCGAGCTGACGGGACCCAGCTCAAGGAG 8861
 QY 282 SerLeuileLysAlaAlaLeuileAlaGlyAlaThrAspValGlyLeuGlyTyProSer 301
 DB 8862 GCATGGTCAAGGAGCAAGCAAGCGCGG-----TACACCCCG 8903
 QY 302 GlyAspGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaValValAsn 321
 DB 8904 TACCAGCGCGCGCGCGCGCTGACGCGCGCGCGCGCTGACACACCGGCTTCGCC 8963
 QY 322 GluAlaThrAlaLeuThrThrGlyGlnLysAlaThrTySerPheGlnThrGlnAlaGly 341
 DB 8964 ACCACGACCGCTACTCC-----GGCTTCCACAG----- 8993
 QY 342 LysProLeuLysIleSerLeuValTrpThrAspAlaProGlySerThr-----Thr 358
 DB 8994 -----TGCCCCCGAAGCCCGGGAGACCGATGTCGGGACG 9029
 QY 359 AlaSerTyThrLeuValAsnAsp-----LeuAspLeuValile-----Thr 372
 DB 9030 GTGACGTACCAACGCTCGCGAGCGCCCGGTGACGCTCAACCTGGCGTCAAGCGCAC 9089
 QY 373 AlaProAsnGlyGlnLysTyThrValGlyAsnAspPheSerTyProTyAspAsnAsnTrp 392
 DB 9090 GTCCCGCGCGGTG-----TTGAGCTCTCCGAGGACCAT----- 9125
 QY 393 AspGlyArgAsnValGluAsnValPheileAsnAlaProGlnSerGlyThrTyThr 412
 DB 9126 -----GTACCGTCCCGCGCGCACCGGACCGCCAGC 9155
 QY 413 IleGluVal---GlnAlaTyAsnValProSerGlyProGlnArgPheSerLeuAla 431
 DB 9156 GTACCGTCCCGCGCGCTGGACAAGCTGGCGGCGGACCGTCCGTCAGCGCGCTGATC 9215

RESULT 6

US-09-974-300-1934
 ; Sequence 1934, Application US/09974300
 ; Patent No. US20020146721A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Berka, Randy M.
 ; APPLICANT: Clausen, Ib Groth
 ; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
 ; TITLE OF INVENTION: Expression
 ; FILE REFERENCES: 10085.500-US
 ; CURRENT APPLICATION NUMBER: US/09/974,300
 ; CURRENT FILING DATE: 2001-10-05
 ; PRIOR APPLICATION NUMBER: 09/680,598
 ; PRIOR FILING DATE: 2000-10-06
 ; PRIOR APPLICATION NUMBER: 60/279,526
 ; PRIOR FILING DATE: 2001-03-27
 ; NUMBER OF SEQ ID NOS: 8481
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 1934
 ; LENGTH: 1329
 ; TYPE: DNA
 ; ORGANISM: Bacillus licheniformis
 US-09-974-300-1934

Alignment Scores:

Pred. No.: 8.58e-27 Length: 1329
 Score: 336.00 Matches: 105
 Percent Similarity: 44.58% Conservative: 43
 Best Local Similarity: 31.63% Mismatches: 116
 Query Match: 14.85% Indels: 68
 DB: 9 Gaps: 12

US-09-985-689a-3 (1-433) x US-09-974-300-1934 (1-1329)

QY 11 AspValAlaGlnAsnAsnTyGlyLeuTyGlyGlnGlyGlnValValAlaValAlaAsp 30
 DB 406 GAAGTGGTCAGAAACAATCAGACCGTCAAGGCAAGAGGTGACAGTCGCTGCTTGTAT 465

QY 31 ThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGluAlaPheArgGlyLysIle 50
 DB 466 ACCGGGGTA-----TACCCTCAGCAAGATCTTGAAGCGCAGATC 504
 QY 51 ThrAlaLeuTyR---AlaLeuGlyArgThrAsnAsnAlaAsnAspProAsnGlyHisGly 69
 DB 505 AGGCGCTTTTCAAGACTTTTATCACCAGAGAACACACCCCTATGATGACATGGGCGCGC 564
 QY 70 ThrHisValAlaGlySerValLeuGlyAsnAlaLeu-----AsnLysGly 84
 DB 565 ACACATGCGCGCGTGTGCTTCGGAACGAGCGCGCTCATCGGGTCAGTACCGCGGA 624
 QY 85 MetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSerSerGlyGlyLeuGly 104
 DB 625 CTGCTCTCTCAGCAGAACTTGTGCGTGTAAAGTATTGGACAAAATGGATCC---GGA 681
 QY 105 GlyLeu-----Pro 107
 DB 682 TCGCTCGAAACCGCTCATTCAGCGGTAGATTGGTGCATTCAATTCATAAGGAAATCCT 741
 QY 108 SerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsnAlaGlyAlaArgIleHisThrAsn 127
 DB 742 GATGATCCGATCGACATTATTCAATGTTCATTTGGTGCAGAGCCCTTGGCTACGAGAAT 801
 QY 128 SerTrpGlyAlaPro-----ValAsnGlyAlaTyThrAlaAsnSerArgGln 143
 DB 802 GAAGAAAGATCCAGTCTTAAGCTGTTTCGACGATGGGACGCGC----- 852
 QY 144 ValAspGluTyThrValArgAsnAsnAspMetThrValLeuPheAlaAlaGlyAsnGluGly 163
 DB 853 -----ATTGTTGTATGTGCGGCGAGCGCAACTCCGCT 885
 QY 164 ProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAla 183
 DB 886 CCGTCCGCAACGATGCGCGCGGTGTCGACGAGCAAGATTATTACAGTCGAGGCC 945
 QY 184 ThrGluAsnTyArgProSerPheGlySerIleAlaAspAsnProAsnHisIleAlaGln 203
 DB 946 TTGGATGACAGG-----GATACAGTCAGCGGAGGATGACGATGTGCGCTCT 993
 QY 204 PheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspValThrAlaProGly 223
 DB 994 TATTCAAGCAGAGCGCGCAATCTATGTCAGTCAACCGGACTTGTGTGTCGCGGC 1053
 QY 224 ThrPheileLeuSerAlaArgSer-----SerLeuAlaProAspSerSerPheTrpAla 241
 DB 1054 ACAATATTACGTCGCTCGTTTCCCGGATCTTTCTCGATAAGCTGCAAAAACAAAC 1113
 QY 242 AsnTyAsnSerLysTyAlaTyMetGlyThrSerMetAlaThrProIleValAla 261
 DB 1114 AGAGTCGCGCAAAAATATGACATTGTCGGAACCTCGATGGCTACCGCATCTGCGCA 1173
 QY 262 GlyAsnValAlaGlnLeuArgGluHisPheileLysAsnArgGlyIleThrProLysPro 281
 DB 1174 GGAATGCGCA-----CTTATCTTCAGCAAGCGCGCGGCACAGAACCT 1218
 QY 282 SerLeuileLysAlaAlaLeuileAlaGlyAlaThrAspVal----- 295
 DB 1219 GATGAAGTCAACAGCTGCTAATGGACGCT---ACCGATTATGGAAGATCGCGATCCA 1275
 QY 296 -----GlyLeuGlyTyProSerGlyAspGln 304
 DB 1276 AATGTTTACGTGCGGATCATCAACGCAAGAACAA 1311

RESULT 7

US-10-156-761-5701
 ; Sequence 5701, Application US/10156761
 ; Publication No. US20030119018A1
 ; GENERAL INFORMATION:
 ; APPLICANT: OMURA, SATOSHI
 ; APPLICANT: IKEDA, HARUO
 ; APPLICANT: ISHIKAWA, JUN

```
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIORITY FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 5701
; LENGTH: 3624
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(3624)
US-10-156-761-5701

Alignment Scores:
Pred. No.: 2,158-25 Length: 3624
Score: 329.00 Matches: 119
Percent Similarity: 42.75% Conservative: 49
Best Local Similarity: 30.28% Mismatches: 161
Query Match: 14.54% Indels: 64
DB: 14 Gaps: 16

US-09-985-689A-3 (1-433) x US-10-156-761-5701 (1-3624)

QY 8 ValLysAlaAspValAlaGlnAsn-----Tyr 17
556 GTGAGGCGGACATGCGCGGAGGACGCGGAGATCGGTACGGCGCGCGCGCGGTGGAGAGGCC 615
18 GlyLeuTyrGlyGlnGlyGlnValValAlaValAlaLeuThrGlyLeuAspThrGlyArg 37
616 GGGCTCACGGGCGGCGGTGTCACCGTCCGCGGTGTCGACACCGCGGTGCGACCC----- 669

QY 38 AsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGly 57
670 -----ACTCACCGCGACCTCGCGCGCGGTGTCGCGGAGCAAGAGCTTCATC 717

QY 58 ArgThrAsnAsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeu 77
718 GACGGGAGGAGGTGCGCGGACCGGACCGGACCGGACCGGACCGGACCGGACCGGACCGG 777

QY 78 Gly-----AsnAlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuVal 92
778 GCGAGCGCGCGCGCTCCGACGCGGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 837

QY 93 PheGlnSerIleMetAspSerSerGlyGlyLeuGlyLeuProSerAsnLeuAsnThr 112
838 GTCGCGCAAGTCTCAGCGGACCGGACCGGACCGGACCGGACCGGACCGGACCGGACCGG 885

QY 113 LeuPheSerGlnAlaThrAsnAla-----GlyAlaArgIleHisThrAsnSerTyr 129
886 ATCGGGGCGATGATGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 945

QY 130 GlyAla---ProValAsnGlyAlaTyrThrAlaAsnSerArgGlnValAspGluTyrVal 148
946 GGATCGACCGAGCGGACCGGACCGGACCGGACCGGACCGGACCGGACCGGACCGGACCGG 1005

QY 149 ArgAsnAsnAspMetThrValLeuPheAlaAlaGlyAsnGluGlyProAsnSerGlyThr 168
1006 GAGGAGACCGGCGCGCTCTGCTGCGCGCGGCGGAGAACCGGCGGCGGCGGCGGCGGCGG 1062

QY 169 IleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnTyrArg 188
1063 ATCGGCTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1107

QY 189 ProSerPheGlySerIleAlaAspAsnProAsnHisIleAlaGlnPheSerSerArgGly 208
```

```
Db 1108 -----GTGACTCATCCGACCGCGCGCGCTACTTACACGACCGCGGC 1149
QY 209 AlaThrArg---AspGlyArgIleLysProAspValThrAlaProGlyThrPheIleLeu 227
1150 CCGCGCACCGGCGCAACCGCTCAAGCGCGACCTCGCGGCGACCGCGGCGTCCGACATCCGC 1209
QY 228 SerAlaArgSerSerLeuAlaProAspSerSerPheTyrAlaAsnTyrAsnSerLysTyr 247
1210 GCGGCGCGCTCCAGCTCGCGCGCGCGCGCGCTAC-----TAC 1248
QY 248 AlaTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeu 267
1249 ACCTCCATGAGCGGTACGTGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 1308
QY 268 ArgGluHisPheIleLysAsnArgGlyIleThrProLysProSerLeuIleLysAlaAla 287
1309 GCGGAGCAGCAGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1368
QY 288 -----LeuIleAlaGlyAlaThrAspValGlyLeuGlyTyrProSerGlyAspGlnGly 305
1369 GAGCAACTCGACGCTCCGTATATCAGTTGGGGGGGGTGGGTGAGTGTGCGGACGCC 1428
QY 306 TrpGly---ArgValThrLeuAspLysSerLeuAsnValAlaTyr-----Valaen 321
1429 GTCGCGCGCGGTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1488
QY 322 GluAlaThrAlaLeuThrThrGlyGlnLysAlaThrTyrSerPheGlnThrGlnAlaGly 341
1489 GACGCGGATCGACCGCGTCAAG-----AAGACGCGTCACTTCCAACTCTCCGAC 1539
QY 342 LysProLeuLysIleSerLeuValTyrThrAspAlaProGlySerThrThrAlaSerTyr 361
1540 ACAGCGGTGAGTTGAGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1590
QY 362 ThrLeuValAsnAspLeuAspLeuValIleThrAlaPro 374
1591 ACCTCGCC-----GACACCGCACCTACCGTGGCC 1620

RESULT 8
US-10-156-761-1
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIORITY FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Alignment Scores:
Pred. No.: 1,798-20 Length: 9025608
Score: 329.00 Matches: 119
```

```

Percent Similarity:   42.75%      Conservative:    49
Best Local Similarity: 30.26%      Mismatches:     161
Query Match:         14.54%       Indels:        64
DB:                  14           Gaps:          16

US-09-985-685A-3 (1-433) X US-10-156-761-1 (1-9025608)

QY      8 ValLeuAlaAspValAlaGlnAsn-----Tyr 17
Db      6918813 GTCCAGGGCGCATTCCTCCTGCCTGTACATTTCAAACTCTGCCGCTTGCCGACTGCGGCGGC 6918872

QY      18 GlyLeuTyrglyglnvalValalalaaspThrHisValAlaGlyserValleu 37
Db      6918873 GGCGCTCACGGGACGGCTCACCGTGGCGTGTCTGACACCGGCGTGCACACC----- 6918926

QY      38 AsnAspSerMetHisGlualPheArgGlyLysIleThraLaLeuTyralaleugly 57
Db      6918927 -----ACTCACCCCAGCTTCCGCGCGCGGGTGTCCGAGGAAGAAGTTTATT 6918974

QY      58 ArgThrAsnAlaAsnAspProAsnglHisglyThrHisValAlaGlyserval 77
Db      6918975 GACGGGAGGAGGTCCGACCGAACCGCCACGGACCCAGCTCACCTCAGCGGTGGC 6919034

QY      78 Gly-----AsnalaeuenLyseGlyMetAlaproglinalasnuVal 92
Db      6919035 GGAGCGGCCCCCTCCAGCGCACGAGCGCGCGTGTCCCGCGTCCAGCTCGCC 6919094

QY      93 PheGlnSeriMetAspSerSerglyLeuGlyLeuProSerAsnLeuasnThr 112
Db      6919095 GTCCGCAAAGTGTACGACGACGAGCGCG-----GGAAGCGAGTCCCAGATC 6919142

QY      113 LeuPheSerGlnatrpasna-----GlylaargilehisThrAsnSertip 129
Db      6919143 ATCCGGGCGATGAATGGCGCGCGGAGCTGCTGCGACGATGCTCTCGTAGAGCTC 6919202

QY      130 GlyAla---ProValansgilyalytThrTralaasnSeraGlnValaspGluTyrrVal 148
Db      6919203 GAATCGACCGAGGCGACGAGCGGACCGACCCCATCGCGCGCGCTCGACACCTCTCC 6919262

QY      149 ArgAsnAspMetThrValLeuPheAlaalaglysnsluclyProAsnsrGlyThr 168
Db      6919263 GAGAGACCGCGGCGCTTCCTGCTGCTGCGCGCGGAAAACCGGTGCCCTCC----TCG 6919319

QY      169 IleSerAlaproglyThrAlaLysAsnalailethValiglyAlathrGluasntyrrArg 188
Db      6919320 ATCCGCTCGCGCGCGCGCGCGAGCTCGCGCTCAGCGTGGCGCC----- 6919364

QY      189 ProSerPheGlyserlleAlaaspAsnpProAsnhisllaaglInPheSerSerrglyly 208
Db      6919365 -----GTCGACTATCGACCGCGCGCGCTTACTTACAGCGCGCGC 6919406

QY      209 AlaThrArg---AspglyArghilleypProspValthrAlaproglyThrPhelleLeu 227
Db      6919407 CCGCGCCAACGCGACAACCGCTCAAGCGGACCTCGCGCACCGCGCGCTCGACATCCGC 6919466

QY      228 SerAlaArgSerSerLeualaProCasPaserPhetrpAlaasntyrserytyr 247
Db      6919467 CGCGCGCGCTCCAGCTCGCGCGCGCGCGCGCTTCTGCGCGCGCGCGCTTC 6919505

QY      248 AlatyrmetylglyThrsermetalthrProlleValalglyasnVallaginLeu 267
Db      6919506 ACCTCCATGAGCGGTACGTGATGCGACGCGCGCATGTCGCGGGGTGCGCGCGCTTCC 6919565

QY      268 ArgGluHissPheIlleYsaasnarGlytleThrpolyssprosrerleulleysAlaA 287
Db      6919565 GCCGAGACGACCCCGAGTGAACCGCGCGCGGTCTGAGACGGGTGATGTCACCGTCC 6919625

QY      288 ----LeuilleAlagiAlathrasppvalcglyLeuglyTyrrProserglyAspGlngly 305
Db      6919626 GAGCAACTCGAGCCCTCATATACGTGCGGCGCGGTCTCGGTTCAGTGTGCGCGAGCGC 6919685

QY      306 TrpGly--ArgvalithrLeuasplyserserleuanValalatyr-----Valasn 321

```

RESULT 9

| | |
|---|--|
| US-10-156-761-3306 Sequence 3306, Application US/10156761 Publication No. US20030119018A1 GENERAL INFORMATION: APPLICANT: OMURA, SATOSHI APPLICANT: IKEDA, HARUO APPLICANT: ISHIKAWA, JUN APPLICANT: HORIKAWA, HIROSHI APPLICANT: SHIBA, TADAYOSHI APPLICANT: SAKAKI, YOSHIOKI APPLICANT: HATTORI, MASAHIRA TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES FILE REFERENCE: 249-262 CURRENT APPLICATION NUMBER: US/10/156, 761 CURRENT FILING DATE: 2002-05-29 PRIOR APPLICATION NUMBER: JP 2001-204089 PRIOR FILING DATE: 2001-05-30 PRIOR APPLICATION NUMBER: JP 2001-272697 PRIOR FILING DATE: 2001-08-02 NUMBER OF SEQ ID NOS: 15109 SEQ ID NO 3306 LENGTH: 3417 TYPE: DNA ORGANISM: Streptomyces avermitilis FEATURE: NAME/KEY: CDS LOCATION: (1)..(3417) US-10-156-761-3306 | Alignment Scores: Pred. No.: 3,27e+25 Length: 3417 Score: 327.00 Matches: 114 Percent Similarity: 45.27% Conservative: 44 Best Local Similarity: 32.66% Mismatches: 126 Query Match: 14.45% Indels: 67 DB: Gaps: 12 |
|---|--|

US-09-985-689A-3 (1-433) x US-10-156-761-3306 (1-3417)

| | |
|---|--|
| QY 18 GlyLeuTyrglyglnGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArg 37 Db 745 GGGTATCAGCGCAAGGCGCTCAAGATCGCGCTCGTGGACACCGGTGTCGAC----- 795 | QY 38 AsnAspSerMetHisGlualPheArgGlyLysIleThraLaLeuTyrrAlaLeugly 57 Db 796 -----CGRACCCACCGGACCTCAGGACCGAGTGGCGGCTCCCAAGACTTCTCC 846 |
| QY 58 ArgThrAsnAlaAsnApproAsnglyHisglyThrHisValalaglyserValleu 77 Db 847 GCCTCCGCGCGCGCGACCACTTCGGTCTACGGCAGCACGCTCGCTCCATCCGGCGG 906 | QY 78 Gly-----AsnalaeuenLyseGlyMetAlaproglinalaasnLeuVal 92 Db 907 GGCACCGCGCGCGCGAGTCCACGCGCAGTACAGGCTGTGCGCGCGCGCGCGATCTCTC 966 |
| QY 93 PheGlnSeriMetAspSerSergly-----GlyLeu 103 Db 957 AACGGCAAGTCTCTCGACGACACCGCTCCGCGCAGATCTCCGCGCATCTCCGCGCGCATG 1026 | |

```
QY 104 -GlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsnAlaGlyAlaAr 123
Db 1027 GAGTGGCGCGCGAGCAGCGCGCGAGCGTCTCAACTGAGCTGGCGCGCGAGCAC 1086
QY 123 gileHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTrpThrAlaAsnSerArgGl 143
Db 1087 CCGAGATCGACCGCGTGGAA-GCGAGGTCAACAAGCTC----- 1125
QY 143 nValAspGluTrpValArgAsnAsnAspMetThrValLeuPhe-----AlaAlaGlyAs 161
Db 1126 -----TCCGAGGAGAGAGGCGATCTCTTCGCGATCGCGCGCGCAA 1166
QY 161 nGluGlyPro---AsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleTh 180
Db 1167 CGAAGGCGAGTTCGGCGAGCAGACATCGGTCTCCCGGCGAGCGCGCGCGCGCTCAC 1226
QY 180 rValGlyAlaThrGluAsnTyArgProSerPheGlySerIleAlaAspAsnProAsnHi 200
Db 1227 CGTCGGCGCC-----GTGAACGACGAGCAA 1253
QY 200 sileAlaGlnPheSerArgGlyAlaThrArgAspGlyArgIleLysProAspValTh 220
Db 1254 GCTGGCGTCTTCCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1313
QY 220 rAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTr 240
Db 1314 CGCACCGCGCGTGGACATCACCGCGCC-----GCGCGCGCGCGCGCGCGCGCGCG 1364
QY 240 pAlaAsnTyArgSerLys-----TyAlaTrpMetGlyGlyThrSerMetAlaTh 257
Db 1365 CCAGAGGTCTCGCGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1424
QY 257 rProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyI 277
Db 1425 CCGCATGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1467
QY 277 eThrProLysProSerLeu-----IleLysAlaAlaLeuIleAlaGlyAlaThrAspVa 295
Db 1468 -----CCCAACTGGTCTGGTCCGCGAGCTCAAGCGCGCGCGCGCGCGCGCGCG 1518
QY 295 lGlyLeuGlyTyProSerGlyAspGlnGlyTrpGlyArgValThrLeuAspLysSerLe 315
Db 1519 -GCGCGGAGTACACCGCGTTCAGCAGGCGTCCGCGCGCGCGCGCGCGCGCGCGCGCG 1577
QY 315 uAsnValAlaTyValAsnGluAlaThrAlaLeuThrThrGlyGlnLysAlaThrTySe 335
Db 1578 CAAGCAGTCCGTGATCGCCCAACCGCAACTCGGTGAGCTTCGGCAT-CCAGCAGTGGCGC 1636
QY 335 rPheGlnThrGlnAlaGlyLysPro 343
Db 1637 ACACGAGACAAAGCGGTCAACCA 1661

RESULT 10
US-10-156-761-1/c
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
```

```
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
; US-10-156-761-1

Alignment Scores:
Pred. No.: 2,96e-20 Length: 9025608
Score: 327.00 Matches: 114
Percent Similarity: 45.2% Conservative: 44
Best Local Similarity: 32.6% Mismatches: 126
Query Match: 14.4% Indels: 67
DB: 14 Gaps: 12

US-09-985-689A-3 (1-433) x US-10-156-761-1 (1-9025608)

QY 18 GlyLeuTyGlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArg 37
Db 4132505 GGGTACGACGGCAAGCGGTCAAGATCGCGTCTCGGACACCGCGTGCAC----- 4132455
QY 38 AsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyArgAlaLeuGly 57
Db 4132454 -----GCGACCCACCGCGACCTCAAGGACCAAGTGGCGCGAGTCCCAAGAACTTCTCC 4132404
QY 58 ArgThrAsnAsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeu 77
Db 4132403 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4132344
QY 78 Gly-----AsnAlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuVal 92
Db 4132343 GGCACCGCGCGCAAGTCCAGCGCAAGTACAGGTTGTGCGCGCGCGCGCGCGCGCG 4132284
QY 93 PheGlnSerIleMetAspSerSerGly-----GlyLeu 103
Db 4132283 AACGGCAAGTCTCTCGACACACCGCGTCCGCGCGAGCTCGCGCATCTCGCGCGCATG 4132224
QY 104 -GlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsnAlaGlyAlaAr 123
Db 4132223 GAGTGGCGCGCGAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4132164
QY 123 gileHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyThrAlaAsnSerArgGl 143
Db 4132163 CCGAGATCGACCGCGTGGAA-GCGAGGTCAACAAGCTC----- 4132125
QY 143 nValAspGluTyValArgAsnAsnAspMetThrValLeuPhe-----AlaAlaGlyAs 161
Db 4132124 -----TCCGAGGAGAGGCGATCTCTTCGCGATCGCGCGCGCAA 4132084
QY 161 nGluGlyPro---AsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleTh 180
Db 4132083 CGAAGGCGAGTTCGGCGAGCAGACCATCGGTCTCCCGGCGAGCGCGCGCGCGCGCGCG 4132024
QY 180 rValGlyAlaThrGluAsnTyArgProSerPheGlySerIleAlaAspAsnProAsnHi 200
Db 4132023 CGTCGGCGCC-----GTGAACGACGAGCAA 4131997
QY 200 sileAlaGlnPheSerArgGlyAlaThrArgAspGlyArgIleLysProAspValTh 220
Db 4131996 GCTGGCGTCTCTTCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4131937
QY 220 rAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTr 240
Db 4131936 CGCACCGCGGTGGACATCACCGCGCC-----GCGCGCGCGCGCGCGCGCGCGCG 4131886
QY 240 pAlaAsnTyArgSerLys-----TyAlaTyMetGlyGlyThrSerMetAlaTh 257
Db 4131885 CCAGGAGTTCGGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4131826
```

```
QY 257 rProfileValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyI1 277
Db 4131825 CCGCATGTCGCGGGCGCGCGGATCTCAAGCAGCAGCAC----- 4131783
QY 277 eThrProLysProSerLeu-----IleLysAlaIleLeuIleAlaGlyAlaThrAspVa 295
Db 4131782 ----CCCAACTGCTGTCGCGAGCTCAAGGCGCGCTGACCGCGCTCGCGAAG----- 4131732
QY 295 lGlyLeuGlyTyProSerGlyAspGlnGlyTyProGlyArgValThrLeuAspLysSerLe 315
Db 4131731 -GGCGCAAGTACACCGCTTCAGCAGGCGCTCGGCGCTATCGCGCGTGCAGACGCGAT 4131673
QY 315 uAsnValAlaTyValAsnGluAlaThrAlaLeuThrThrGlyGlnLysAlaThrTySe 335
Db 4131672 CACGACGTCGTCGTCGCGCAACCGAAGTCTCGTGAGCTTCGGCAT-CCAGCAGTGGCGGC 4131614
QY 335 rPheGlnThrGlnAlaGlyLysPro 343
Db 4131613 ACACCGACGACGACGCGGTCAACCA 4131589

RESULT 11
US-10-090-624-5
; Sequence 5, Application US/10090624
; Publication No. US20020132335A1
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOUO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA=6
; CURRENT APPLICATION NUMBER: US/10/090,624
; CURRENT FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: 09/445,472
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 5
; LENGTH: 4765
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-10-090-624-5

Alignment Scores:
Pred. No.: 1,37e-22 Length: 4765
Score: 305.00 Matches: 128
Percent Similarity: 36.26% Conservative: 37
Best Local Similarity: 28.13% Mismatches: 154
Query Match: 13.48% Indels: 136
DB: 13 Gaps: 15

US-09-985-689a-3 (1-433) x US-10-090-624-5 (1-4765)
QY 21 GlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 1024 GGCAATGTTATGACATTTGCATATGTGATGATGACCTTGACCTGACCTACCGACGAA 1083
QY 41 -----SerMetHisGluAlaPheArgGly 48
Db 1084 GTTCCACTTGGCCAGTCAACGTTACTTATGATGTTGCTGTTTTAGTACTACTACGGT 1143
QY 49 LysIleThrAlaLeuLeuLeuLeuGlyArgThrAsnAsnAlaAsnAspProAsn----- 66
Db 1144 CCTCTCAAC-----TACGTGCTTCAGAAATA-----GATCCTCAACGGAGAA 1185
QY 67 -----GlyHisGlyThrHisValAlaGlySerValLeu 77
Db 1186 TATGCAGTATTGGGTGGGTGGTGCACGGTCAACGAACTCAGGTAGTGGAACTGTGCT 1245
```

```
QY 78 GlyAsnAlaLeuAsn----- 82
Db 1246 GGTTACGACGACCAACAATGATGCTTGGGATTGGCTCAGTATGATCTCTGTGTAATGGAA 1305
QY 83 -----LysGly 84
Db 1306 GTGTTCTCAAGACTCTATGTTGGGATTTATCAAGAGTTTACCACAGACCCGTGAGGGT 1365
QY 85 MetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSerSerGlyGlyLeuGly 104
Db 1366 GTTGCTCCAGGTGCCCAATAATAGCAATAAGAGTTCTTAGGAGTGTATGGA-----CGG 1419
QY 105 GlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsnAlaGlyAlaAArgIle 124
Db 1420 GGTAAGCATGTGGGATATTAGAAAGGTATGACATACGCAACCAACCCATGTGTGACAGCT 1479
QY 125 HisThrAsnSerTrpGly-----AlaProValAsnGlyAlaTyThrAlaAsnSerArg 142
Db 1480 ATAAGCATGAGTCTCGTGGGAATGCTCCATCTTAGTGGTACTGATCCAGAAAGCGTT 1539
QY 143 GlnValAspGluTyValArgAsnAsnAspMetThrValLeuPheAlaAlaGlyAsnGlu 162
Db 1540 GCTGTGATGAGCTTACCGAAAAGTACGGTGTGTATTCGTATATAGTGTGAGGAATGAA 1599
QY 163 GlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGly 182
Db 1600 GGTCTGCGCATTAACATCGTTGGAGTCTCTGGTGTGCAACAAGGCAATACTGTGGA 1659
QY 183 AlaThrGlu----- 185
Db 1660 GCTGCTGAGTGGCCCATTAACGTTGGAGTTATGTTTCCCAAGCACTTGGATATCTCGAT 1719
QY 186 -----AsnTyArgProSerPheGlySerIleAlaAspAsnProAsnHisIle 201
Db 1720 TACTATGAGTCTTATTACTCCCGCGCTACACAACGTT-----AGAATA 1764
QY 202 AlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspValThrAla 221
Db 1765 GCATTCTCTCAGCAGAGGCGCGAGAATAGATGTTGAAATAAAACCAATAGTAGTGGCT 1824
QY 222 ProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAla 241
Db 1825 CCAGTTACGGAATTTACTATCCCTCCGATGGGATGGGAGTGGCGGAGCTGACTTC----- 1878
QY 242 AsnTyArgSerLysTyArgAlaTyMetGlyGlyThrSerMetAlaThrProIleValAla 261
Db 1879 -----ATGCTCGAACTTCGATGGCTACTCCACATGTCCAGC 1914
QY 262 GlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrProLysPro 281
Db 1915 GGTGTCGTGTGACCTCCCTCATAGCGGG---GCAAGGCGCGAGGAGATATACTACAATCCA 1971
QY 282 SerLeuIleLysAlaAlaLeuIleAlaGlyAlaThr-----AspVal 295
Db 1972 GATATAATTAAGAAGTTCTTGAGCGGTGCAACCTGGCTTGAGGAGATCCATATACT 2031
QY 296 GlyLeuGlyTyProSerGlyAspGlnGlyTrpGlyArgValThrLeuAspLysSerLeu 315
Db 2032 GGGCAGAAAGTACACTGAGCTTGACCAAGTCTCATGCTTGTTAACGTTACCAAGTCTG 2091
QY 316 AsnValAlaTyArgValAsnGluAlaThrAlaLeuThrThrGlyGlnLysAlaThrTySer 335
Db 2092 GAAATC----- 2097
QY 336 PheGlnThrGlnAlaGlyLysProLeuLysIleSerLeuValThrThrAspAlaProGly 355
Db 2098 CTTAAGGCTATAAAGCGCACCACTCTCCCAATTTGATCACTCGGAGAGCAAGTCTTAC 2157
QY 356 SerThrThrAlaSerTyThrLeuValAsnAspLeuValIleThrAlaProAsn 375
Db 2158 AGCGACTTTGGGAGTACTTGGGTGTG---GACGTTATAAGAGGTCTCTACGCAAGGAC 2214
```



```
; PRIOR APPLICATION NUMBER: PCT/EP01/09815
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: DE 101 09 166.4
; PRIOR FILING DATE: 2001-02-25
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn Ver. 3.2
; SEQ ID NO 113
; LENGTH: 1560
; TYPE: DNA
; ORGANISM: Streptomyces viridochromogenes
; FEATURE:
; OTHER INFORMATION: avx16 dna: partial sequence of coding strand 1; nucleotide 1
; OTHER INFORMATION: corresponds to nucleotide 56,198 of coding strand 1.
US-10-084-846A-113

Alignment Scores:
Pred. No.: 3,39e-22 Length: 1560
Score: 295.00 Matches: 125
Percent Similarity: 42.11% Conservative: 51
Best Local Similarity: 29.90% Mismatches: 153
Query Match: 13.04% Indels: 90
DB: 15 Gaps: 19

US-09-985-689A-3 (1-433) x US-10-084-846A-113 (1-1560)
Qy 21 GlyGlnGlyValValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAspSer 40
Db 451 GGGCAGGAGTACGGGTACGTCATCGACACCGGGTCTC-----CGC 492
Qy 41 SerMetHisGluAlaPheAeGGVLYsalleThrAlaLeuTyrr---AlaLeuGlyArgThr 59
Db 493 ATCACCCACAGCGATTCTGGCGCGCGCTCTCTAGCGGTACGACGCCATCGACACGAC 552
Qy 60 AsnAsnAlaAsnAspProAnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn 79
Db 553 AACACCCCGCCAGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 612
Qy 80 AlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 99
Db 613 GCCTTAC-----GGCGTCGCAAGAGGCAAGGCGCGTCTAGCGTCCGCGTGTGAACAAC 666
Qy 100 Ser-----GlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAla 117
Db 667 TCCGGCCAGGCGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 726
Qy 118 TrpAsnAlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaPro-ValAsnGlyAlaTy 137
Db 727 GT-CAAGCGCGCGTCCGCAACAT-----GTCCCTCGCGCGCGCGCGCGCGCGCGCG 770
Qy 137 rThrAlaAsnSerArgGlnValAlaAspGluTyValArgAsn-----AsnAspMetTh 154
Db 771 CACGGCC-----CTCGACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 818
Qy 154 rValLeuPheAlaAlaGlyAsnGlyProAsnSerGlyThrIleSerAlaProGlyTh 174
Db 819 CTTGCGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 875
Qy 174 rAlaLysAsnAlaIleThrValGlyAlaThrGluAsnTyArgProSerPheGlySerIl 194
Db 876 GGTACACCGAGGCCATCAAGTCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 930
Qy 194 eAlaAsnAsnProAsnHisIleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyAr 214
Db 931 -----TATCCAACTACGCTCTCCGCTCTCTC----- 954
Qy 214 gIleLysProAspValThrAlaProGlyThrPheIleLeuSerAlaArgSerLeuAl 234
Db 955 -----GACCTCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 990
Qy 234 aProAspSerSerPheTrpAlaAsnTyArgAsnSerLysTyAlaTyMetGlyGlyThrSe 254
Db 991 -----TGGAACTCAAGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1034
```

```
254 rMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAs 274
1035 GATGGCAGCCCGCAGCGTGGCGGGCGCGCGCGCGCTC-----CACCTCCGCCCA 1085
Qy 274 rArgGlyIleThrProLys-----ProSerLeuIleLysAlaAlaLeuIleAlaGl 291
Db 1086 CCCCTCGGCCACCGCGTCCCGAGTCCCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1145
Qy 291 yAlaThrAspValGlyLeuGlyTyrrProSerGlyAspGlnGlyTrpGlyArgValThrLe 311
Db 1146 CGTCACCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1176
Qy 311 uAspLysSerLeuAsnValAlaTyValAsnGluAlaThrAlaLeuThrThrGlyGlnLy 331
Db 1177 -----CGGCTCTGTACGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1223
Qy 331 s-----AlaThrTyrrSerPheGlnThrGlnAlaGlyLysProLeuLy 345
Db 1224 CTTGAGAGAACACCGGTGACTACAGATCAGCGACAACTCCACCGTGGAGTCCCGCGTGAC 1283
Qy 345 sIleSerLeuValTrpThrAspAlaProGlySerThrThrAlaSer-----Ty 361
Db 1284 GGTCTCGGGCGTCTCCCGGCAACCGCGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1343
Qy 361 rThrLeuValAsnAspLeuValIleThrAlaProAsnGlyGlnLysTyrrValGl 381
Db 1344 CACGTACATCGCGACCTCCAGGTCCAGTGTATCGCCCCCGACGCGCACGGCGTACACGCT 1403
Qy 381 yAsnAspPheSerTyrrProTyrrAspAsnAsnTrpAspGlyArgAsnAsnValGluAsnVa 401
Db 1404 CAAG-----TCGTACGCGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 1457
Qy 401 lPheIleAsnAlaProGln-----SerGlyThrTyrrThrIleGluVal 415
Db 1458 G-----AACGCTCTCTCGAGGCGCGCAACGCGCACGCTGGAAACTGCGGGGTG 1503

RESULT 14
US-10-084-846A-1
; Sequence 1, Application US/10084846A
; Publication No. US20040006026A1
; GENERAL INFORMATION:
; APPLICANT: WEITNAUER, GABRIELE
; APPLICANT: MUHLNWEIG, AGNES
; APPLICANT: TREFFZER, AXEL
; APPLICANT: BECHTOLD, ANDREAS
; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES
; FILE REFERENCE: 1974-005
; CURRENT APPLICATION NUMBER: US/10/084,846A
; CURRENT FILING DATE: 2003-02-25
; PRIOR APPLICATION NUMBER: PCT/EP01/09815
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: DE 101 09 166.4
; PRIOR FILING DATE: 2001-02-25
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn Ver. 3.2
; SEQ ID NO 1
; LENGTH: 59816
; TYPE: DNA
; ORGANISM: Streptomyces viridochromogenes
US-10-084-846A-1

Alignment Scores:
Pred. No.: 6,68e-20 Length: 59816
Score: 295.00 Matches: 125
Percent Similarity: 42.11% Conservative: 51
Best Local Similarity: 29.90% Mismatches: 153
Query Match: 13.04% Indels: 90
DB: 15 Gaps: 19

US-09-985-689A-3 (1-433) x US-10-084-846A-1 (1-59816)
Qy 21 GlyGlnGlyValValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAspSer 40
```

56648 GGGCAGGAGTACGGCGTACGTATCATCGACACCGCGGTC-----CGC 56689
Qy 41 SerMetHisGluAlaPheAArgGlyLeuValAlaLeuTyr---AlaLeuGlyArgThr 59
Db 56690 ATCACCACACGCGCTTCGGCGGCGGCGCTTCAGCGCTACGACGCCATCGAACACGAC 56749
Qy 60 AsnAsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn 79
Db 56750 AACACCGCCAGAGCGCCACGGCCAGCGACGACGTCGTGCCGCGACGTCGCCGCGAAC 56809
Qy 80 AlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerileMetAspSer 99
Db 56810 GCCTAC-----GGCGTCGCCAAGAGCCCAAGATCGTAGGCGTCGGCGGTGAACAAC 56863
Qy 100 Ser-----GlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAla 117
Db 56864 TCCGGCAGGCGACACCGCGCGAGTCGTCCGGCGATCGACTGGGTGCGCCGGACGCGC 56923
Qy 118 TrpAsnAlaGlyAlaArgileHisThrAsnSerTrpGlyAlaPro-ValAsnGlyAlaTy 137
Db 56924 GT-CAAGCGCGCGTCGCCAACAT-----GTCCCTCGCGCGCGCGCGA 56967
Qy 137 rThrAlaAsnSerArgGlnValAspGlyTyrValArgAsn-----AsnAspMetTh 154
Db 56968 CACGGCC-----CTCGACACGGCGGTACGACCAAGCCATCGATGGCTCCGGCGGTAC 57015
Qy 154 rValLeuPheAlaAlaGlyAsnGlyProAsnSerGlyThrileSerAlaProGlyTh 174
Db 57016 CTTCCGCGTGGCGCGCGACGAGTCGACCAACGCGCTCCACGAGGTCA---CCGCGACG 57072
Qy 174 rAlaLysAsnAlaLeuThrValGlyAlaThrGluAsnTyrArgProSerPheGlySerIl 194
Db 57073 CGTCACGAGGCCATCAGTCGGCGCGCGACGACGCTCGGACGCCCAAGCCCGCGC----- 57127
Qy 194 eAlaAspAsnProAsnHisAlaGlnPheSerSerArgGlyAlaThrArgAspGlyAr 214
Db 57128 -----TACTCAACTACGGCTCCGCTC----- 57151
Qy 214 glleLysProAspValThrAlaProGlyThrPheileLeuSerAlaArgSerleuAl 234
Db 57152 -----GACCTCTTCGCGCGCGGTCTGCTCATCATCCTCGGCC----- 57187
Qy 234 aProAspSerPheTtAlaAsnTyrAsnSerLysTyrAlaTyrMetGlyGlyThrSe 254
Db 57188 -----TGAAGTCAAGGACTCGCGGACCAACACCATCTCCGGTAGCTC 57231
Qy 254 rMetAlaThrProfileValAlaGlyAsnValAlaGlnLeuArgGluHisPheileLysAs 274
Db 57232 GATGGCGACCGCGACGTCGGCGCGCGCGCGCTC-----CACCTCGCGCGCAA 57282
Qy 274 nArgGlyileThrProLys-----ProSerLeuileLysAlaAlaLeuileAlaGl 291
Db 57283 CCGCTCGGCGCCCGCTCCAGGTCCGACGCGCTGACGTCCGCGCGCCACCGCGCGT 57342
Qy 291 yAlaThrAspValGlyLeuGlyTyrProSerGlyAspGlnGlyTrpGlyArgValThrLe 311
Db 57343 CGTCACCAACCGCGCGACGGGCTCGCCCAAC----- 57373
Qy 311 uAspLysSerLeuAsnValAlaTyrValAsnGlnAlaThrAlaLeuThrThrGlyGlnLy 331
Db 57374 -----CGGCTCTGTAGTCTCGCGCGCGCGACCAACCGCTCCGGCGCGCG 57420
Qy 331 s-----AlaThrTyrSerPheGlnThrGlnAlaGlyLysProLeuLy 345
Db 57421 CTTTCGAGAACCGGTGACTACGATCAGCGACAACTCCACGGTCGAGTCCCGGTGAC 57480
Qy 345 sileSerLeuValTrpThrAspAlaProGlySerThrThrAlaSer-----Ty 361
Db 57481 GGTCTTCGGCGGTCTCCGGCAACGCGCGCTCCGCGCTCCGCGTAGAGGTCCACATCGTCCA 57540
Qy 361 rThrLeuValAsnAspLeuAspLeuValileThrAlaProAsnGlyGlnLysTyrValGl 381
Db 57541 CACGTACATCGCGGACCTTCAGGTCCAGTGTATCGCCCCCGACCGCGCGGTACCGCT 57600

Qy 381 yAsnAspPheSerTyrProTyrAspAsnAsnTrpAspGlyArgAsnAsnValGluAsnVa 401
Db 57601 CAAG-----TCGTACGGCACCGCGCGAGTTCGGACAAATCAACACCATCTCGT 57654
Qy 401 lPheileAsnAlaProGln-----SerGlyThrTyrThrileGluVal 415
Db 57655 G-----AACGCTCTCTCGGAGCGCGCAACGCGCACGTGGAACTCGCGGTG 57700

RESULT 15

US-10-084-846A-2/c
; Sequence 2, Application US/10084846A
; Publication No. US20040006026A1
; GENERAL INFORMATION:
; APPLICANT: WEITNAUER, GABRIELE
; APPLICANT: MUHLNENEG, AGNES
; APPLICANT: TREFFZER, AXEL
; APPLICANT: BECHTHOLD, ANDREAS
; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES
; FILE REFERENCE: 1974-005
; CURRENT APPLICATION NUMBER: US/10/084,846A
; CURRENT FILING DATE: 2003-02-25
; PRIOR APPLICATION NUMBER: PCT/EP01/09815
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: DE 101 09 166.4
; PRIOR FILING DATE: 2001-02-25
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn Ver. 3.2
; SEQ ID NO 2
; TYPE: DNA
; ORGANISM: Streptomyces viridochromogenes
US-10-084-846A-2

Alignment Scores:
Pred. No.: 6,68e-20 Length: 59816
Score: 295.00 Matches: 125
Percent Similarity: 42.11% Conservative: 51
Best Local Similarity: 25.90% Mismatches: 153
Query Match: 13.04% Indels: 90
DB: 15 Gaps: 19

US-09-985-689A-3 (1-433) x US-10-084-846A-2 (1-59816)

Qy 21 GlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 3169 GGGCAGGAGTACCGCGTACGTATCGACACCGCGCTC-----CGC 3128
Qy 41 SerMetHisGluAlaPheAArgGlyLysileThrAlaLeuTyr---AlaLeuGlyArgThr 59
Db 3127 ATCACCACACGCGACTTCGGCGCGCGGCTCTCTACGCTACGACGCTACGACCAACGAC 3068
Qy 60 AsnAsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn 79
Db 3067 AACACCGCCAGAGCGCGCCACCGCGCACGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3008
Qy 80 AlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerileMetAspSer 99
Db 3007 GCCTAC-----GGCGTCGCCAAGAGCCCAAGATCGTAGGCGTCGCGGTGCTGAACAAC 2954
Qy 100 Ser-----GlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAla 117
Db 2953 TCCGGCG 2894
Qy 118 TrpAsnAlaGlyAlaArgileHisThrAsnSerTrpGlyAlaPro-ValAsnGlyAlaTy 137
Db 2893 GT-CAAGCGCGCGCTCGCCAACAT-----GTCCCTCGCGCGCGCGCGCGCG 2850
Qy 137 rThrAlaAsnSerArgGlnValAspGlyTyrValArgAsn-----AsnAspMetTh 154
Db 2849 CACGGCC-----CTCGACACGGCGGTACGCAACGCGCATGGCTCCGGCGGTAC 2802
Qy 154 rValLeuPheAlaAlaGlyAsnGlyProAsnSerGlyThrileSerAlaProGlyTh 174


```
Db 2801 CTTCCGCGTGGCGCGCAACGAGTCGACCAACGCTCCACGAGGTCA---CCGCGACG 2745
Qy 174 rAlaLysAsnAlaIleThrValGlyAlaThrGluAsnTyrArgProSerPheGlySerIle 194
Db 2744 CGTCACCGAGCGCATCAGGTCCGCGCGACGACCGCTCGGACGCCAACGCGCGC----- 2690
Qy 194 eAlaAspAsnProAsnHisIleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArg 214
Db 2689 -----TACTCCAACTACGGCTCCGTCCTC----- 2666
Qy 214 gIleLysProAspValThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAl 234
Db 2665 -----GACCTCTTCGCCCCCGGTTCGTCATCACCCTCGGCC----- 2630
Qy 234 aProAspSerPheThrAlaAsnTyrAsnSerLysTyrAlaTyrMetGlyGlyThrSe 254
Db 2629 -----TGGAACTCAAGGACTCGCGGACCAACACCATCTCCGGTACGTC 2586
Qy 254 rMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAs 274
Db 2585 GATGGCAGACCCCGACGTCGCGCGCGCGCGCGCTC-----CACCTCGCGCCCAA 2535
Qy 274 nArgGlyIleThrProLys-----ProSerLeuIleLysAlaAlaLeuIleAlaG1 291
Db 2534 CCCCTCGGCCACCCCGTCCCGAGTCGCCACGGGCTGACGTCGCGCGCCACCAACCGCGGT 2475
Qy 291 yAlaThrAspValGlyLeuGlyTyrProSerGlyAspGlnGlyTyrGlyArgValThrLe 311
Db 2474 CGTCACCAACCCCGGACGCGGCTCGGCCAAC----- 2444
Qy 311 uAspLysSerLeuAsnValAlaTyrValAsnGluAlaThrAlaLeuThrThrGlyGlnLy 331
Db 2443 -----CGGCTCCTGTACGTCGCGCGCGCGGACGACCACTCCGCGGCGCGCG 2397
Qy 331 s-----AlaThrTyrSerPheGlnThrGlnAlaGlyLysProLeuLy 345
Db 2396 CTTGGAACACCGGTACTACGATCAGATCAGCAACTCCACGTCGAGTCCCGGTGAC 2337
Qy 345 sIleSerLeuValThrPheAlaProGlySerThrThrAlaSer-----Ty 361
Db 2336 GGTCTCCGGGTCTCCGCGCAACGCGCTCCGCGCTCGGCGCTCCACATCGTCCA 2277
Qy 361 rThrLeuValAsnAspLeuValIleThrAlaProAsnGlyGlnLysTyrValG1 381
Db 2276 CACGTATATCGCGACCTCCAGTTCAGTTCAGTTCGCGCGCGCGCGCGGTACACGT 2217
Qy 381 yAsnAspPheSerTyrProTyrAspAsnAsnThrPheGlyArgAsnAsnValGluAsnVa 401
Db 2216 CAAG-----TCGTACGGCACCGCGCGAGTTCGGACAAACATCAACACCGCTACTCGGT 2163
Qy 401 lPheIleAsnAlaProGln-----SerGlyThrTyrThrIleGluVal 415
Db 2162 G-----AACCCCTCTCGGAGCGCGCAACGCGACGTGGAACTCGCGGTG 2117
```

Search completed: April 4, 2004, 21:12:07
Job time : 6907.25 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 3, 2004, 23:40:05 ; Search time 2274.58 Seconds
(without alignments)
5684.703 Million cell updates/sec

Title: US-09-985-689A-3
Perfect score: 2263
Sequence: 1 NDVARGIVKADVAQNNYGLY.....EVQAYNVSPGQRFSLAIVH 433

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 27513289 seqs, 14931090276 residues
Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2.1/USPTO.spool/US09985689/runat.31032004.161807.4168/app_query.fasta_1.3498
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09985689@cgn.1.1.7084/runat.31032004.161807.4168 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPOP=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*

29: gb_gss2:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|-------|-------------|--------|----|-----------|---------------------------------------|
| 1 | 235.5 | 10.4 | 1605 | 13 | BO622771 | CC Contig BO622771 CC Contig |
| 2 | 228 | 10.1 | 4198 | 11 | AK029048 | Mus muscu AK029048 Mus muscu |
| 3 | 214.5 | 9.5 | 640 | 12 | BJ395336 | Bj395336 Bj395336 |
| 4 | 213.5 | 9.4 | 594 | 12 | BQ393752 | Bj393752 Bj393752 |
| c | 204 | 9.0 | 2141 | 13 | BQ142519 | Contig96 M BQ142519 Contig96 M |
| 6 | 200.5 | 8.9 | 532 | 29 | TA319G10P | AL494464 T. brucei AL494464 T. brucei |
| 7 | 199.5 | 8.8 | 718 | 12 | BI750157 | Fg02_1090 BI750157 Fg02_1090 |
| 8 | 195.5 | 8.6 | 771 | 14 | CA320325 | UI-M-FW0- CA320325 UI-M-FW0- |
| 9 | 193.5 | 8.6 | 601 | 12 | Bj387574 | Bj387574 Bj387574 |
| 10 | 190 | 8.4 | 545 | 13 | Bj575479 | TgBST2v8 Bj575479 TgBST2v8 |
| 11 | 188 | 8.3 | 574 | 29 | TA319H10P | AL490202 T. brucei AL490202 T. brucei |
| 12 | 188 | 8.3 | 614 | 9 | AJ273402 | AJ273402 AJ273402 |
| c | 185.5 | 8.2 | 2121 | 28 | B2424995 | B2424995 100023066 B2424995 100023066 |
| 14 | 182.5 | 8.1 | 665 | 13 | BQ770462 | UI-M-F10- BQ770462 UI-M-F10- |
| c | 182.5 | 8.1 | 716 | 28 | B2893395 | B2893395 HL2_0177 B2893395 HL2_0177 |
| c | 180 | 8.0 | 530 | 29 | CNS010PO | AL153820 Anopheles AL153820 Anopheles |
| 17 | 180 | 8.0 | 650 | 9 | AJ274038 | AJ274038 AJ274038 |
| c | 179.5 | 7.9 | 675 | 14 | CF728224 | UI-M-HB0- CF728224 UI-M-HB0- |
| 18 | 179 | 7.9 | 508 | 28 | AQ652212 | Sheared D AQ652212 Sheared D |
| c | 178 | 7.9 | 633 | 12 | Bj369190 | Bj369190 Bj369190 |
| 20 | 177.5 | 7.8 | 576 | 14 | CD295943 | StrPu691. CD295943 StrPu691. |
| 22 | 177.5 | 7.8 | 1002 | 29 | CNS06D6B | CF846026 psHB036XA CF846026 psHB036XA |
| 23 | 176.5 | 7.8 | 594 | 14 | CF846026 | T7 end of CF846026 T7 end of |
| c | 176.5 | 7.8 | 1029 | 29 | CNS071DW | AJ273745 AJ273745 |
| 25 | 175 | 7.7 | 580 | 9 | AJ273947 | AJ273947 AJ273947 |
| 26 | 175 | 7.7 | 583 | 9 | AJ273918 | AJ273918 AJ273918 |
| 27 | 175 | 7.7 | 593 | 9 | AJ273050 | AJ273050 AJ273050 |
| 28 | 175 | 7.7 | 601 | 9 | AJ273921 | AJ273921 AJ273921 |
| 29 | 175 | 7.7 | 601 | 12 | Bj393925 | Bj393925 Bj393925 |
| 30 | 173.5 | 7.7 | 641 | 12 | AJ274218 | AJ274218 AJ274218 |
| 31 | 172 | 7.6 | 610 | 28 | AQ651427 | Sheared D AQ651427 Sheared D |
| c | 170 | 7.5 | 584 | 28 | AQ651427 | UI-M-EW0- AQ651427 UI-M-EW0- |
| 33 | 169.5 | 7.5 | 781 | 13 | BQ612128 | BQ612128 BQ612128 |
| 34 | 169.5 | 7.5 | 937 | 12 | BG246418 | BG246418 BG246418 |
| 35 | 169 | 7.5 | 609 | 9 | AJ272712 | AJ272712 AJ272712 |
| 36 | 169 | 7.5 | 794 | 14 | CF737198 | CF737198 CF737198 |
| 37 | 167.5 | 7.4 | 604 | 9 | AJ274126 | AJ274126 AJ274126 |
| 38 | 167.5 | 7.4 | 615 | 9 | AJ274059 | AJ274059 AJ274059 |
| 39 | 167 | 7.4 | 616 | 14 | CF138007 | CF138007 CF138007 |
| 40 | 167 | 7.4 | 681 | 14 | CB690041 | CB690041 CB690041 |
| 41 | 166.5 | 7.4 | 895 | 13 | BQ216158 | BQ216158 BQ216158 |
| 42 | 166 | 7.3 | 619 | 9 | AJ273097 | AJ273097 AJ273097 |
| 43 | 166 | 7.3 | 887 | 13 | BQ879057 | BQ879057 BQ879057 |
| 44 | 165.5 | 7.3 | 593 | 9 | AJ273903 | AJ273903 AJ273903 |
| 45 | 165.5 | 7.3 | 596 | 9 | AJ273379 | AJ273379 AJ273379 |

ALIGNMENTS

RESULT 1
BO622771
LOCUS BO622771
DEFINITION CC Contig967 Conidiobolus cornatus ARSEF 512 Conidiobolus coronatus
ACCESSION BO622771
VERSION BO622771.1
KEYWORDS GI:21649940
SOURCE Conidiobolus coronatus (Delacroixia coronata)
ORGANISM Conidiobolus coronatus
Eukaryota; Fungi; Zygomycota; Zygomycetes; Entomophthorales; Ancylistaceae; Conidiobolus.
REFERENCE 1 (bases 1 to 1605)

BO622771 1605 bp mRNA linear EST 01-JUL-2002
CC Contig967 Conidiobolus cornatus ARSEF 512 Conidiobolus coronatus
cDNA, mRNA sequence.
BO622771
BO622771.1 GI:21649940
EST.
Conidiobolus coronatus (Delacroixia coronata)
Conidiobolus coronatus
Eukaryota; Fungi; Zygomycota; Zygomycetes; Entomophthorales; Ancylistaceae; Conidiobolus.
1 (bases 1 to 1605)

Db 1201 GCTGCTGGTAACGATAACAGAGATGCTTGTGGATACTCT---CCGCGTTCTGCTCCTTCC 125


```

QY 314 SerLeuAsnValAlaTyrValAsnGluAlaThrAlaLeuThrThrGlyGlnLysAlaThr 333
   :::
Db 1827 GCTTATCAGATC-----CTC 1841

QY 334 TyrSerPheGlnThrGlnAlaGly---LysProLeuLysIleSerLeu----- 348
   |||
Db 1842 AGCAGCTATAACCCGAGCGAGCTGAGTCTAGCTACATCGACCTGAGTGTGCC 1901

QY 349 ---ValTrpThrAspAlaProGlySerThrThrAlaSerTyr-----Thr 362
   |||
Db 1902 TACATGTGG-----CCCTACTGCTCCGACGCTATCTACTATGGAGGAATGCCAAC 1952

QY 363 LeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyGlnLysTyrValGlyAsn 382
   |||
Db 1953 ATCGTTAT-----GTCACCATCCTCAATGCGATGGCGGTCCACAGGAAGA 1997

QY 383 AspPheSerTyrProTyrAspAsnAsnTrp-----AspGlyValAsnAsn 397
   |||
Db 1998 ATTGTGGATAGCCT-----GAGTGGCGACCCCTATTTACCAACAGATGGAGACAA 2048

QY 398 ValGluAsnValPhe-----IleAsnAlaProGlnSerGlyThrTyrThrIle 413
   |||
Db 2049 ATTGAAGTGGCTTCTCCTACTCCTCAGTGTGTGGCCCTGGTCAGGTACCTTGGCATC 2108

QY 414 GluVal 415
   ::
Db 2109 TCATT 2114

RESULT 3
LOCUS BJ395336
DEFINITION BJ395336 Dictyostelium discoideum cDNA library, SF Dictyostelium
discoideum cDNA clone dds32b16 5', mRNA sequence.
ACCESSION BJ395336.1 GI:19306422
VERSION BJ395336
KEYWORDS EST.
SOURCE Dictyostelium discoideum
ORGANISM Dictyostelium discoideum
REFERENCE 1 (bases 1 to 640)
AUTHORS Eukaryota; Mycetozoa; Dictyostelidia; Dictyostelium.
TITLES Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
JOURNAL Full length cDNA of Dictyostelium discoideum at the slug stage
COMMENT Unpublished (2002)
Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.

FEATURES
     source
     1..640
     /organism="Dictyostelium discoideum"
     /mol_type="mRNA"
     /strain="AX4"
     /db_xref="taxon:44689"
     /clone="dds32b16"
     /sex="mat A"
     /dev_stage="slug stage"
     /clone_lib="Dictyostelium discoideum cDNA library, SF"

ORIGIN
Alignment Scores:
Pred. No.: 2,24e-11 Length: 640
Score: 214.50 Matches: 67
Percent Similarity: 50.47% Conservative: 41
Best Local Similarity: 31.31% Mismatches: 65
Query Match: 9.48% Indels: 41
DB: 12 Gaps: 12

US-09-985-689A-3 (1-433) x BJ395336 (1-640)

```

```

QY 19 LeuTyrGlyGlnGlyGlnValValAlaAspThrGlyLeuAspThrGlyArg--- 37
   |||
Db 22 TTAAGAGGTAAAGGTCAGATATTGAGTATTGCTGATCTGTTAGATGGTAGCCATTGT 81

QY 38 ---AsnAspSer-----SerMetHisGluAlaPheArgGly 48
   |||
Db 82 TTCTTTTCAGATCAAGTATCCATATCAATTTATCAAGTGAATGAATAATCATAGAAA 141

QY 49 LysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAsnAspPro-----Asn 66
   |||
Db 142 GTTGTAACTTATATT-----ACTTACCATGACAATGAAGATTATGTAAAT 186

QY 67 GlyHisGlyThrHisValalaglySerValLeuGlyAsn-----Ala 80
   |||
Db 187 GGTCAATGGTACACATTTTGTGGCTCTGCAGCAGGTACTCCAGAGATTTCTCATGGGCT 246

QY 81 LeuAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAsp 98
   |||
Db 247 ATTTCATCATTTAGTGTCTTCACTGACGCAAGATTGCTATTTATGATCTT----- 300

QY 99 SerSerGlyGlyLeuGlyLeu---ProSerAsnLeuAsnThrLeuPheSerGlnAla 117
   |||
Db 301 TCATCTGGAAGTTCTGAACCAACACACCCGAGATTACAGTCAATGTACAAACCATTA 360

QY 118 TrpAsnAlaGlyAlaArgIleHisThrAsnSerTrpGlyAla-----ProVal 133
   |||
Db 361 TATGATGCAGGTCCAGAGGTACATGCTGATTTCTTGGGTTCTGTATCTTTGCCAAGTTAT 420

QY 134 AsnGlyAlaTyrThrAlaAsnSerArgGlnValAspGluTyrVal---ArgAsnAsnAsp 152
   |||
Db 421 TATGCTGGTTATTCGATGATGCTGGTGTATTTGATGCTATCTCTATGAGTACCCAGAA 480

QY 153 MetThrValLeuPheAlaAlaGlyAsnGluGlyProAsnSerGlyThrIleSerAlaPro 172
   |||
Db 481 TTCTTATATTAAGAGTCTGCTGTAT--AAGAGCTATTTGCACTTTATAGTCAA 537

QY 173 GlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnTyrArgProSerPheGly 192
   |||
Db 538 GCAACAGCTAAAAATGCAATTACAGTTGGTGTGCTGAGCAAAACAGCTCATGTAAATTATGTG 597

QY 193 Ser-----IleAlaAspAsnProAsn 199
   |||
Db 598 TCAGATGCATGGATATTTATGATTCTTCAGATAATGCTAAT 639

RESULT 4
LOCUS BJ393752
DEFINITION BJ393752 Dictyostelium discoideum cDNA library, SF Dictyostelium
discoideum cDNA clone dds32b16 5', mRNA sequence.
ACCESSION BJ393752
VERSION BJ393752.1 GI:19304838
KEYWORDS EST.
SOURCE Dictyostelium discoideum
ORGANISM Dictyostelium discoideum
REFERENCE 1 (bases 1 to 594)
AUTHORS Eukaryota; Mycetozoa; Dictyostelidia; Dictyostelium.
TITLES Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
JOURNAL Full length cDNA of Dictyostelium discoideum at the slug stage
COMMENT Unpublished (2002)
Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.

FEATURES
     source
     1..594
     /organism="Dictyostelium discoideum"
     /mol_type="mRNA"
     /strain="AX4"
     /db_xref="taxon:44689"
     /clone="dds32b16"

```

```

/sex="mat A"
/dev stage="slug stage"
/clone lib=injectovirillum discoideum cdna library. SP"

```

ORIGIN

| | | |
|------------------------|---------|--|
| Alignment Scores: | | |
| Pred. No.: | 2.5e-11 | Length: 594 |
| Score: | 213.50 | Matches: 54 |
| Percent Similarity: | 53.48% | Conservative: 38 |
| Best Local Similarity: | 31.21% | Mismatches: 60 |
| Query Match: | 9.44% | Indels: 21 |
| DB: | 12 | Gaps: 8 |
| | | HIS-09-985-689A-3 (1-433) x BU393752 (1-594) |

```

Qy 255 MetAlaThrProIleValAlaGlyAenValAlaGlnLeuArgGluHisPheIleLeuAsn 274
Db 471 ATGGCTACTCCCATATTGCTGGT-----CTTGCTGCCTACTTCAGTGCTCTC 424
Qy 275 ArgGlyIleThrProLysProSerLeuLeuLeuAlaLeuAlaLeuAlaGlyAlaThrAsp 294
Db 423 ACGCGCAAGACTAGCCCTCCGCTCTTTCCAGAGATCCAGGACACTTCCACCAAGAAC 364
Qy 295 ValGlyLeuGlyTyrProSerGly 302
Db 363 GTGATCCGCAATGTGCCGCTGCG 340

RESULT 6
TA319G10P 532 bp DNA linear GSS 13-DEC-2000
LOCUS T. brucei sheared genomic DNA clone 319g10, forward sequence,
DEFINITION genomic survey sequence.
ACCESSION AL492464
VERSION AL492464.1 GI:11867408
SOURCE Trypanosoma brucei
ORGANISM Trypanosoma brucei
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma.
REFERENCE 1 (bases 1 to 532)
AUTHORS Hall,N., Bowman,S., Lennard,N.J., Doggett,J., Atkin,R.,
Chillingworth,C., Ormond,D., Harris,B., El-Sayed,N., Hou,L.,
Melville,S.E., Rajandream,M.A. and Barrell,B.G.
Direct Submission
Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing
project, Sanger Centre. The Wellcome Trust Genome Campus, Hinxton,
Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and
nh@sanger.ac.uk
COMMENT Constructed at the Institute for Genomic Research (TIGR),
Rockville, MD. Genomic DNA isolated from a cloned population of
Trypanosoma brucei (TREU927/4 GUTAT 10.1) was mechanically sheared
to give a tight size distribution (
4 kb). The v + i method used for the library construction is
described in detail in Smith, H. and Venter, J.C. (Making small
insert libraries for whole genome shotgun sequencing projects. In
Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barrell, Oxford University Press, 1999).
Email: melsayed@tigr.org
Details of T. brucei sequencing at the Sanger Centre are available
at http://www.sanger.ac.uk/projects/T_brucei/.
FEATURES
source
1. 532
/organism="Trypanosoma brucei"
/mol_type="genomic DNA"
/strain="TREU927"
/db_xref="taxon:5691"
/clone="319g10"

ORIGIN
Alignment Scores: 4.45e-10 Length: 532
Pred. No.: 200.50 Matches: 58
Score: 200.50 Conservatve: 27
Percent Similarity: 46.20%
Best Local Similarity: 31.52% Mismatches: 72
Query Match: 8.86% Indels: 27
DB: 29 Gaps: 5

US-09-985-689a-3 (1-433) x TA319G10P (1-532)

Qy 92 ValPheGlnSerIleMetAspSerSerGlyGlyLeuGly----- 104
Db 2 GTAATGCCAAGTATAATGCTGCCAGGCGGCGAAGATTCTTCAGGGGTTGGGCTGCC 61
Qy 105 -----GlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTtp 118
Db 62 CATCCAGTCAGACTTGCTCTCCCCACGACGATTCTCAAAATTATTCGCCGGTATAT 121

```

```

Qy 119 AsnAlaGlyAlaArgIleHisThrAsnSerTtpGlyAlaProValAsnGlyAlaTyrThr 138
Db 122 GGCGCTGGAGCCCGTGTCTCTCAAACTCGTGGGGTTTGTCTCTCCCTCCGAGTATCT 181
Qy 139 AlaAsnSerArgGlnValAspGluTyrValA:GAsn---AsnAspMetThrValLeuPhe 157
Db 182 CTTGTGGAAGAAGATATGATGATGTTGGAGTAGTATGACGATGCGCTACTATCTTC 241
Qy 158 AlaAlaGlyAsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaIleAsn 177
Db 242 TCCACTGGCAACAGTATCCAAAGATGCG-----CTAATGACTCCGCTCGTGGTAAAGAAC 295
Qy 178 AlaIleThrValGlyAlaThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsn 197
Db 296 GTGATGTCGTGGGTCACACAAAAACGTGTTGACGCTTCGAAAGACATT----- 346
Qy 198 ProAsnHisIleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLeuPro 217
Db 347 -----GTTTCTTCGTTTCTTCGCGATGTCCTCAACATACGACGGTAGGATGAACCC 397
Qy 218 AspValThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSer 237
Db 398 GATCTTGTCTCGTCCCGGAGAGGTCGTCTGCTCTTCTCTCGCAAGCATCAGCT 457
Qy 238 SerPheTrpAlaAsnTyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThr 257
Db 458 AAA-----CAATGTAAGTGGTCCCAAGCGGNGTTTCATCGATGGCAACT 502
Qy 258 ProIleValAla 261
Db 503 GCGGCGCTGCGG 514

RESULT 7
BI750157 718 bp mRNA linear EST 25-SEP-2001
LOCUS Gibberella zeae cDNA clone Fg02_10g08, mRNA sequence.
DEFINITION Gibberella zeae cDNA clone Fg02_10g08, mRNA sequence.
ACCESSION BI750157
VERSION BI750157.1 GI:15771959
KEYWORDS EST.
SOURCE Gibberella zeae
ORGANISM Gibberella zeae
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Hypocreomycetidae; Hypocreales; Nectriaceae; Gibberella.
REFERENCE 1 (bases 1 to 718)
AUTHORS Harris,L.J., Glassco,T., Rocheleau,H., Allard,S., Chapados,J.,
Couroux,P., De Moors,A., Hattori,J.I., Ouellet,T., Robert,L.S.,
Singh,J.A, Sprott,D. and Tinker,N.A.
Expressed Sequence Tags from Fusarium graminearum mycelium
Unpublished (2001)
Contact: Harris, Linda J.
Eastern Cereal and Oilseed Research Centre
Agriculture and Agri-food Canada
Bldg. 21, Central Experimental Farm, Ottawa, Ontario, KIA 0C6,
CANADA
Tel: (613) 759-1314
Fax: (613) 759-6566
Email: harris@em.agr.ca
FEATURES
source
1. 718
/organism="Gibberella zeae"
/mol_type="mRNA"
/strain="DAOM 180378"
/db_xref="taxon:5518"
/clone="Fg02_10g08"
/tissue_type="mycelial tissue"
/dev_stages="Asexual"
/clone_lib="Fg02_AAFCC_EOARC_Fusarium_graminearum_mycelium"
/site="Bluescript SK+/XhoI-ECORI; Site 1: EcoRI;
Site 2: XhoI; Mycelial tissue was collected from v8 agar
plates after a growth period of 6-7 days at 25 C with 14
hrs (FL/UV) day lighth exposure. Mycelia was ground in

```

liquid nitrogen prior it's storage at -80 C until RNA extraction. Directional cloning with 5' end of cDNA cloned into EcoRI site of pBluescript and 3' end of cDNA cloned into XhoI site of pBluescript (Stratagene, La Jolla, CA)."

ORIGIN

Alignment Scores:

Pred. No.: 9,44e-10 Length: 718
Score: 199.50 Matches: 70
Percent Similarity: 44.20% Conservative: 29
Best Local Similarity: 31.25% Mismatches: 84
Query Match: 8.82% Indels: 41
DB: 12 Gaps: 9

US-09-985-689A-3 (1-433) x B1750157 (1-718)

Qy 46 PheArgGlyLeuThrAlaLeuTyxAlaLeuGlyArgThrAsnAlaAsnAspPro 65
Db 9 TTCGAGGTCGTCTCAGGCTGTCTACACTGCCCTCAGCGCCAGACGCT--GACACC 65
Qy 66 AsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnAlaLeuAsnLysGlyMet 85
Db 66 AACGGTCACGGAACCTCAGTTGCTGGCACTATTCCCGAAGACATAC-----GGTGT 119
Qy 86 AlaProGlnAlaAsnLeu-----ValPheGln---SerIleMetAspSerSer 100
Db 120 GCCAAGAGGCCACCATCAAGCTGTCCAGGTCCTCCAGGTCATTCATCCAGCACCTCC 179
Qy 101 GlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsnAla 120
Db 180 ATCACTCTCGTGGCTTCACTGGGCTGCCAACGACATCATCTCCAAAG----- 227
Qy 121 GlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTrpThrAlaAsn 140
Db 228 ---GCCGAACCAAGACTCAGTCGTCAATATGCTCTCGCGGTGTTACTCTGCTCTCC 284
Qy 141 SerArgGlnValAspGlyLeuValArgAsnAspMetThrValLeuPheAlaAlaGly 160
Db 285 TTCAACAACGGCTGTCAGTCTKCTCCAGCTCCGGTATTATCTCTGCGCATTTGTCGCGGT 344
Qy 161 AsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThr 180
Db 345 AACGATGTGGCAACGCTGCCAACACTTCT---CTGCTCTTCTCCCGCCGCAWCACT 401
Qy 181 ValGlyAlaThrGluAsnTyxArgProSerPheGlySerIleAlaAspAsnProAsnHis 200
Db 402 GTCGGTGCC-----ATTGACAGCAACTGGGCC 428
Qy 201 IleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspValThr 220
Db 429 ATTCTCTGTACTCCAACTACGTCACCGTTCTC-----GATATCTTT 470
Qy 221 AlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrp 240
Db 471 GCYCTKGACACCGGTTCTCTCCGCC-----TGG 500
Qy 241 AlaAsnTyxAsnSerLysTyxAlaTrpMetGlyThrSerMetAlaThrProIleVal 260
Db 501 TACACCAGCAACAGTKCCACCAACACCATCAGCGGKAGCTCCATGGCTACTCYCCACAT 560
Qy 261 AlaGlyAsnVal 264
Db 561 GCCGGACTGTCTC 572

RESULT 8

CA320325

LOCUS

CA320325

DEFINITION

UI-N-FWO-cby-g-23-0-UI-r1 NIH BMAP_FWO Mus musculus cDNA clone

ACCESSION

CA320325

VERSION

GI:24538449

KEYWORDS

EST.

SOURCE

Mus musculus (house mouse)

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

1 (bases 1 to 771)

AUTHORS

NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE

National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL

Unpublished (1999)

COMMENT

Contact: Robert Strausberg, Ph.D.
Email: cgsbbs-remail.nih.gov
Tissue Procurement: Dr. Jim Lin, University of Iowa
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>

This clone was contributed by the Brain Molecular Anatomy Project (BMAP)

Seq primer: pyx-5.

Location/Qualifiers

1. 771

/organism="Mus musculus"

/mol_type="mRNA"

/strain="CS7BL/6"

/db_xref="taxon:10090"

/clone="IMAGE:6816072"

/tissue_type="whole brain"

/dev_stage="embryo 13.5, 14.5, 15.5, 16.5, 17.5dpc"

/lab_host="DH10B (T1 phage resistant)"

/clone_lib="NIH BMAP_FWO"

/note="Organ: Brain; Vector: pyx-Asc; Site: 1: EcoR I; Site: 2: Not I; The library was constructed according to Bernaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NotI and then cloned directionally into pyx-Asc vector. The library tag sequence located between the Not I site and the polyA tail is AGCAGACAG. This library was created for the University of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemin Chin, Ph.D., Program coordinator."

ORIGIN

Alignment Scores:

Pred. No.: 2,74e-09 Length: 771
Score: 195.50 Matches: 78
Percent Similarity: 40.21% Conservative: 37
Best Local Similarity: 27.27% Mismatches: 114
Query Match: 8.64% Indels: 57
DB: 14 Gaps: 10

US-09-985-689A-3 (1-433) x CA320325 (1-771)

Qy 8 VallyAlaAspValAlaGlnAsnAsnTyxGlyLeuTyxGlyGlnGlyGlnValValAla 27
Db 40 CTCGAGCGCATGTGCTGTGG---CAGATGGGATACACAGGTGCTAATGTCAGATTGCT 96
Qy 28 ValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGluAlaPheArg 47
Db 97 GTTTTGTACTGGCTC-----AGTGAGAAGCATCCGCAATTTTAAG 138
Qy 48 GlyLysIleThrAlaLeuTyxAlaLeuGlyArgThrAsnAsnAlaAsn----- 63
Db 139 AAT-----GTGAGGAGAGAACCACTGGACCAATGAGCGGACCTTG 180
Qy 64 ---AspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnAlaLeuAsn 82
Db 181 GATGATGGGTAGCCCATGCGACATTCGTTGCAAGGTGTGATTGCCACGATGAGGGAGTGC 240


```

QY 83 LysGlyMetAlaProGlnAlaAsnLeu---ValPheGlnSerIleMetAspSerSergly 101
Db 241 CAAGAGATTGCTCCAGATGAGAGCTGCACATCTTCAGGGCTTTACCAACAATCAG--- 297
QY 102 GlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsnAlaGly 121
Db 298 -----GTGCTTTACACACTCTGGTTCTCGGATGCGCTTCAACTATGCC 339
QY 122 AlaAArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTrpThrAlaAsnSer 141
Db 340 ATCTTAAGAAGATGAGACGCTTCAACCTTAGCATCGGTGGCCCGCATTCATCGATCAT 399
QY 142 ArgGlnValAspGluTrp-----ValArgAsnAsnAspMetThrValLeuPheAlaAla 159
Db 400 CGGTTTGTGACAAAGGTGGGAATTAACAGCTAACTAATGTAATATGTTCTGCTATT 459
QY 160 GlyAsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLeuAsnAlaIle 179
Db 460 GGCATGATGAGACCTCTCTATGCACTCTGAATAACCCCTGCTGATCAGATGATGATT 519
QY 180 ThrValGlyAlaThrGluAsnTrpArgProSerPheGlySerIleAlaAspAsnProAsn 199
Db 520 GGAGTGGGTGGC-----ATTGACTTTGAAGAT 546
QY 200 HisIleAlaGlnPheSerSerArgGlyAlaThrArg-----AspGly 213
Db 547 ACATCGCTCGCTTTCTCCAGGGGATGACTACTGGGATTAACAGAGGCTATGCT 606
QY 214 ArgIleLysProAspValThrAlaProGlyThrPheIleLeuSerAlaArgSerLeu 233
Db 607 CGTGTGAAGCCTGACATCTGACCTATGCTGTGGTGGAGTGGGGGT----- 651
QY 234 AlaProAspSerSerPheTrpAlaAsnTrpAsnSerLysTrpAlaTrpMetGlyThr 253
Db 652 -----TCGGTGTGAAGGGGGTGGCTGCGTCACTCTCAGGGACC 690
QY 254 SerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheLeuLys 273
Db 691 AGTGTGCTTCCAGCTGCTGCTGGGGCGTGCACCTTGTAGTANGCAGATACAGAAG 750
QY 274 AsnArgGlyIleThrPro 279
Db 751 CGGAGCTGCTGTAATCCT 768

RESULT 9
BJ387574
LOCUS
DEFINITION
Dictyostelium discoideum cDNA library, SF Dictyostelium
ACCESSION
VERSION
KEYWORDS
SOURCE
Dictyostelium discoideum
ORGANISM
Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
REFERENCE
1. (bases 1 to 601)
Urushihara, H., Tanaka, Y., Kohara, Y. and Shin-i, T.
Full length cDNA of Dictyostelium discoideum at the slug stage
Unpublished (2002)
Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-5856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
Location/Qualifiers
1. .601
/organism="Dictyostelium discoideum"
/mol_type="mRNA"
/strain="AX4"
/db_xref="taxon:44689"
/clones="Gds3a18"

FEATURES
source

```

```

/sex="mat A"
/dev_stage="slug stage"
/clone_lib="Dictyostelium discoideum cDNA library, SF"

```

ORIGIN

```

Alignment Scores:
Pred. No.: 2,87e-09 Length: 601
Score: 139.50 Matches: 55
Percent Similarity: 46.34% Conservative: 21
Best Local Similarity: 33.54% Mismatches: 49
Query Match: 8.55% Indels: 39
DB: 12 Gaps: 6

US-09-985-689A-3 (1-433) x BJ387574 (1-601)
QY 197 AsnProAsnHisIleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLys 216
Db 51 AATGCAATAAATATTTGTCATCTCTCATCAAGGGTCCAAACATCATGCTAGATGAAA 110
QY 217 ProAspValThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAla----- 234
Db 111 CTGATTTAGTTCCTCCCTGCTGTAATATATTACATCGCAAGATCAATGCTGCCAATACA 170
QY 235 -----ProAspSerSerPheTrpAlaAsnTrpAsnSerLysTrpAlaTrpMet 250
Db 171 ACAGACCAATGTGCTGATGCTCTTTA-----CCAAATACAAATGCAATTATTGGCG---ATA 224
QY 251 GlyGlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHis 270
Db 225 TCTGGTACATCAATGGCACTCTTTCGAGCAGCAGCAACCAATCTCTAGACAATAT 284
QY 271 -----PheIleLysAsnArgGlyIleThrProLys 280
Db 285 TTAGTTGATGTTATTATCAACTGTTCAATTGTAGAATCAATAAATTAACAACAAC 344
QY 281 ProSerLeuIleLysAlaAlaLeuIleAlaGlyAla----- 292
Db 345 GGATCATTTTAAAGCAATTAATGATTAATTAATGCTCAGTATTAAATGCTACATTCAA 404
QY 293 -----ThrAspValGlyLeuGlyTyrProSerGlyAsp----- 303
Db 405 TTGATTATCATCATCAAGTATTACATATCATCAACAACCAAGTTTGTGAAATTTTGCAGGT 464
QY 304 -----GlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTrp 319
Db 465 GCAAGTTTAGTTCAAGTTTGGGTGCTATTAGATGAGTAATTGGTTACATGTTGTCAT 524
QY 320 ValAsnGluAla 323
Db 525 AATAATAATAGT 536

RESULT 10
BJ387574
LOCUS
DEFINITION
Dictyostelium discoideum cDNA library, SF Dictyostelium
ACCESSION
VERSION
KEYWORDS
SOURCE
Dictyostelium discoideum
ORGANISM
Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;
Sarcocystidae; Toxoplasma.
1 (bases 1 to 545)
Tang, X., Cole, R., Fogarty, S., Sibley, L.D., Ajioke, J.A., White, M.,
Clifton, S., Fape, D., Martin, J., Wylie, T., Dante, M., Marra, M.,
Hillier, L., Kucaba, T., Theising, B., Bowers, Y., Gibbons, M.,
Ritter, E., Bennett, J., Franklin, C., Tsagareishvili, R., Ronko, I.,
Kennedy, S., Maguire, L., Waterston, R. and Wilson, R.
Toxoplasma EST Project
Unpublished (2001)
Contact: Clifton, S.

```

Toxoplasma EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: toxo@wustl.edu
Contact David Sibley (toxoeast@borcim.wustl.edu) for further
information relating to organism, libraries, or clone availability.
Seq primer: -40RP from Gibco
High quality sequence stop: 445.

Location/Qualifiers
1. 545
/organism="Toxoplasma gondii"
/mol_type="mRNA"
/strain="VEG (Type III)"
/db_xref="taxon:5811"
/clone="TGESrzb80b07.yl"
/dev_stage="Partially sporulated oocysts"
/clone_lib="TgVEG Partially sporulated oocyst cDNA"
/note="Vector: Modified pBluescript (pBSSK+); Site 1:
BamHI; Site 2: EcoRI; PolyA mRNA from partially-sporulated
oocysts was converted to cDNA using the
template-switching PCR method (SMART cDNA, Clontech Inc.,)
and sized selected on SizeSep 400 columns (Amersham
Pharmacia Biotech Inc.). First strand was reverse
transcribed using the CDS III-oligo-dT primer and a 5'
template switch primer (Smart IV primer). The product of
the first strand synthesis was PCR amplified using the
same primer set and the fragments were digested with SfiI.
The fragments were size selected, ligated into a modified
pBluescript vector containing directional SfiI sites, and
electroporated into DH10B or DH12S cells. Vector: SfiI
sites were added to the multiple cloning region of
pBluescript SK+ between the BamHI/EcoRI sites. The
modified polylinker has the following sequence:
5'-GATTCGGCATTCAGCGCC(G)n--insert--
GCCGCCCTCGCCACAGATC3' where n=3-4 G nucleotides. Library
Source: Michael White, Maria E. Jerome, Emily A. Johnson,
Jay A. Radke, Montana State University. Clone
Availability: David Sibley, Washington University"

ORIGIN
Alignment Scores:
Pred. No.: 5,548-09 Length: 545
Score: 190.00 Matches: 64
Percent Similarity: 44.81% Conservative: 31
Best Local Similarity: 30.19% Mismatches: 75
Query Match: 8.40% Indels: 42
DB: 13 Gaps: 9
US-09-985-689A-3 (1-433) x BUS75479 (1-545)
QY 64 AspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGly----- 78
DB 4 GAGCACACGGCCAGCGCATGTGCGCGC---ATCATCGGGCGCATCAAGACCAA 60
QY 79 AsnAlaLeuAsnGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAsp 98
DB 61 AACTCCCGGTGAAGCGGTCTGTGGGAACACGAGCATGCGCGCTTAAGTTCATGGT 120
QY 99 SerSerGlyGlyLeuGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrp 118
DB 121 GCAACAGGA-----AACGGGTGCAGTCAGATGCCATCAAGGCTCTCAACTACGCCGTC 174
QY 119 AsnAlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTrp 138
DB 175 CAGATGGGAATCCCTCAGCTGCACTGCTGGGGCGGTCCACCTGGTCCGAGCGGTG 234
QY 139 AlaAsnSerArgGlnValAspGluTyValArgAsnAsnAspMetThrValLeuPheAla 158
DB 235 ATTGCGGCTCTGGAGCGCGGAGAGCGTGGTGCAC-----CTTTTCATTGCGCG 285
QY 159 AlaGlyAsnGluGlyProAsnSerGlyThrIle-----SerAlaProGlyThrAla 175

Db 286 GCGGAAACCAAGGAAGAACACCGAATTCGGCATTACCTCGCTCGTACCGGCTC 345
QY 176 LysAsnAlaIleThrValGlyAlaThrGluAsnTyArgProSerPheGlySerIleAla 195
Db 346 GCAACAGTAGTCAGCGGTGGTGGACA----- 372
QY 196 AspAsnProAsnHisIleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIle 215
Db 373 AACTCCGAGGACCAACTCGCTTCCTTCAGCAACGGGGAGCGGC----- 417
QY 216 LysProAspValThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaPro 235
Db 418 ACGGTGACCTTGGCGCGCGCGGTAAATCTCTC-----TCTACCTTCGCGCA 468
QY 236 AspSerSerPheTrpAlaAsnTyAsnSerLysTyAlaTrpMetGlyGlyThrSerMet 255
Db 469 GAC-----CAGTTTCGGAACATTTCGGAACCTCCATG 501
QY 256 AlaThrProIleValAlaGlyAsnValAlaGlnLeu 267
Db 502 GCTACGCGCGTGTGCGCGCGGTTCGCGCATCTCTC 537
RESULT 11
TA315H10P 574 bp DNA linear GSS 13-DEC-2000
LOCUS T. brucei sheared genomic DNA clone 315h10, forward sequence.
DEFINITION genomic survey sequence.
ACCESSION AL490202
VERSION AL490202.1 GI:11866292
KEYWORDS GSS.
SOURCE Trypanosoma brucei
ORGANISM Trypanosoma brucei
REFERENCE 1 (bases 1 to 574)
AUTHORS Hall, N., Bowman, S., Lennard, N.J., Doggett, J., Atkin, R.,
Chillingworth, C., Ormond, D., Harris, B., El-Sayed, N., Hou, L.,
Melville, S.E., Rajadream, M.A. and Barrell, B.G.
TITLE Direct Submissiion
JOURNAL Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing
project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,
Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and
nhl@sanger.ac.uk
COMMENT Constructed at the Institute for Genomic Research (TIGR),
Rockville, MD. Genomic DNA isolated from a cloned population of
Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared
to give a tight size distribution (4 kb). The v + i method used for the library construction is
described in detail in Smith, H. and Venter, J.C. (Making small
insert libraries for whole genome shotgun sequencing projects. In
Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barrell, Oxford University Press, 1999).
Email: released@tigr.org
Details of T. brucei sequencing at the Sanger Centre are available
at http://www.sanger.ac.uk/projects/T_brucei/
Location/Qualifiers
1. 574
/organism="Trypanosoma brucei"
/mol_type="genomic DNA"
/strain="TREU927"
/db_xref="taxon:5691"
/clone="315h10"

ORIGIN
Alignment Scores:
Pred. No.: 9,72e-09 Length: 574
Score: 188.00 Matches: 51
Percent Similarity: 47.65% Conservative: 30
Best Local Similarity: 30.00% Mismatches: 59
Query Match: 8.31% Indels: 30
DB: 29 Gaps: 6